Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **449,** 3082–3106 (2015) doi:10.1093/mnras/stv463

On optical mass estimation methods for galaxy groups

R. J. Pearson,[1⋆] T. J. Ponman,[1⋆] P. Norberg,[2] A. S. G. Robotham[3] and W. M. Farr[1]

[1]*School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*
[2]*ICC, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[3]*International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*

Accepted 2015 March 2. Received 2015 March 1; in original form 2014 August 30

ABSTRACT

We examine the performance of a variety of different estimators for the mass of galaxy groups, based on their galaxy distribution alone. We draw galaxies from the Sloan Digital Sky Survey for a set of groups and clusters for which hydrostatic mass estimates based on high-quality *Chandra* X-ray data are available. These are used to calibrate the galaxy-based mass proxies, and to test their performance. Richness, luminosity, galaxy overdensity, rms radius and dynamical mass proxies are all explored. These different mass indicators all have their merits, and we argue that using them in combination can provide protection against being misled by the effects of dynamical disturbance or variations in star formation efficiency. Using them in this way leads us to infer the presence of significant non-statistical scatter in the X-ray based mass estimates we employ. We apply a similar analysis to a set of mock groups derived from applying a semi-analytic galaxy formation code to the Millennium dark matter simulation. The relations between halo mass and the mass proxies differ significantly in some cases from those seen in the observational groups, and we discuss possible reasons for this.

Key words: galaxies: groups: general.

1 INTRODUCTION

An understanding of the mass of galaxy groups and clusters is essential to the study of these structures and their evolution. For example, use of the concept of self-similarity, whereby larger systems behave as scaled up versions of smaller ones (e.g. Kaiser 1986; Navarro, Frenk & White 1997; Alard 2013), requires knowledge of the mass of the objects in question. Studies of self-similarity based on cosmological simulations have direct access to dark matter particle information, which in turn allows the user to construct robustly defined halo masses. However, this luxury is not available in the case of observational studies, and comparisons of baryonic properties, such as X-ray luminosity and gas entropy, with self-similar expectations require reliable estimates of the mass of the host haloes.

In practice, groups and clusters are observed to depart from self-similar expectations (Arnaud & Evrard 1999; Ponman, Cannon & Navarro 1999; Neumann & Arnaud 2001; Osmond & Ponman 2004), and these departures offer valuable evidence about the nature of additional astrophysical processes, such as cooling and cosmic feedback, which break the symmetries observed in simple gravity-only models. Estimates of halo mass, and the corresponding characteristic overdensity radii of systems, are an essential element in the study of the baryon astrophysics. The effects of these additional processes are especially notable within poor galaxy groups ($M \lesssim 10^{14}$ M$_\odot$), so it is unfortunate that it is precisely in such systems that masses are most difficult to determine.

Most existing methods of mass estimation rely on the group or cluster mass distribution affecting a baryonic tracer population, which then provides a measurable mass proxy. The simplest, and oldest, of these approaches uses the dynamics of member galaxies within the cluster halo, whose velocity dispersion is related to system mass through the Virial Theorem (e.g. Zwicky 1937). This approach is often applied to large redshift surveys as a useful and relatively straightforward mass estimator (e.g. Eke et al. 2004; Ramella et al. 2004; Robotham et al. 2011). This estimator requires that group membership be well-established, with limited contamination from foreground and background galaxies, though robust estimators of group velocity dispersion, such as the *gapper* estimator (Beers, Flynn & Gebhardt 1990), help to reduce the impact of outliers in the galaxy velocity distribution. The method also requires that the galaxy tracers are relaxed – i.e. their motions are not strongly affected by dynamical disturbances such as group mergers or infall. Studies of simulated clusters by Biviano et al. (2006) indicate that virial mass estimates are subject to biases of 10 per cent or more. There are also indications, from comparisons between galaxy dynamics and X-ray temperatures, that in poor groups velocity dispersions may be subject to unexplained downward biases which can be substantially larger (Osmond & Ponman 2004).

An alternative approach uses the hot gas within the dark matter halo of groups and clusters. This gas, heated by gravitational collapse to virial temperatures, radiates X-rays. The

⋆ E-mail: richard@star.sr.bham.ac.uk (RJP); tjp@star.sr.bham.ac.uk (TJP)

surface brightness distribution and spectrum of this radiation can be used to infer the radial distribution of gas density and temperature, and hence, via hydrostatic equilibrium, to estimate system mass (e.g. Mulchaey 2000). Such an analysis requires high-quality X-ray data, with radially resolved spectroscopy. With poorer quality X-ray data, it is still possible to derive useful mass estimates if a mean X-ray temperature can be measured, applying well-established mass–temperature relations for groups and clusters (e.g. Sun et al. 2009). However, even this may not be available for many galaxy groups, since they tend to have low X-ray surface brightness. Moreover, as X-ray data are expensive to acquire and existing all sky surveys such as RASS (*ROSAT* All Sky Survey; Voges et al. 1999) are too shallow on average, high-quality X-ray spectral imaging is still not available for the majority of the sky. The upcoming surveys to be performed by the *eROSITA* instrument (Merloni et al. 2012), may help moderate this, though the detection of the lowest surface brightness groups may still be a challenge.

The most direct estimator of gravitating mass is gravitational lensing. The distortions imposed on the light from background galaxies as it passes through the gravitational potential of a foreground cluster allows the surface density of mass along the line of sight to be derived (Narayan & Bartelmann 1996). However, the magnitude of these lensing distortions is so small for poor groups that at present lensing studies at such low masses are largely confined to stacking analyses. These extract an average mass distribution, destroying information on the properties of individual groups (Parker et al. 2005).

Due to the availability of large galaxy surveys, the ability to estimate the masses of galaxy groups and clusters from their galaxy contents alone is an attractive possibility. The aim of this paper is therefore to explore ways of estimating masses given only basic properties of member galaxies available from surveys such as SDSS (Ahn et al. 2014) and GAMA (Driver et al. 2011). We will evaluate the performance of these estimators using a sample of groups for which X-ray-based masses have been well-constrained.

X-ray-bright groups appear to constitute a particular subset of the group population (e.g. Rasmussen et al. 2006), so to check these results and obtain access to a wider range of groups, we will also investigate the performance of our mass proxies on a sample of groups drawn from a cosmological simulation.

In Section 2, we discuss the data used in this study. Section 3 presents the mass proxies that we use and in Section 5, we present the results for each methodology. In Section 6 and Section 7, we present and discuss analogous results for a sample of groups drawn from a cosmological simulation. Finally, in Section 8, we discuss the implications of our results for the practical problem of estimating the masses of galaxy groups.

Throughout the paper, we adopt a simple Λ cold dark matter cosmology with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$. When converting literature results to our cosmology, we adopt $h = H_0/100\,{\rm km\,s^{-1}\,Mpc^{-1}} = 0.7$ when required. Due to the small range in current values of H_0 and the relatively large scatter in mass proxies observed, the choice of cosmology is not a significant concern; we therefore report values in physical units rather than as a function of h_{70}, where $h_{70} = H_0/70\,{\rm km\,s^{-1}\,Mpc^{-1}}$. The analysis performed here makes use of the R statistical package (R Development Core Team 2009).[1]

[1] www.r-project.org

2 DATA

This study requires a sample of galaxy groups with both optical data, for use in constructing mass proxies, and robust mass estimates that are independent of optical properties, against which mass–proxy scaling relations can be investigated and calibrated. We take our canonical masses from high-quality X-ray analyses and use the Sloan Digital Sky Survey (SDSS) as the source of our optical data.

2.1 X-ray

Our galaxy group sample is primarily drawn from the Sun et al. (2009, hereafter S09) *Chandra* study of galaxy groups. This uses a sample of 43 groups with Advanced CCD Imaging Spectrometer (ACIS) observations from the *Chandra* archive. The groups were selected to have low temperature ($T_{500} \lesssim 2.7\,{\rm keV}$) and redshift ($0.015 < z < 0.13$), together with a relaxed morphology (i.e. emission is not significantly substructured or disturbed). These criteria ensure that these groups have low mass and high-quality X-ray data suitable for a hydrostatic mass analysis. The lower redshift limit ensures that r_{2500}, the radius enclosing a mean density 2500 times the critical density of the Universe, for each group lies within the ACIS field of view, allowing S09 to trace gas properties to large radii.

The result is one of the most robust X-ray analyses of low mass groups currently available. Additionally, the low redshifts of these groups implies that their member galaxies should be well sampled by the SDSS for groups which lie within the Sloan survey area. It should be noted that this sample is an X-ray-selected sample and may differ systematically from optically selected samples, as we will discuss later.

The groups of S09 were split into four tiers depending on the extent to which they were able to trace gas properties from the emission centre. We use the two best subsets for which M_{500} was either measured at r_{500} or was extrapolated based on gas properties at a large fraction of r_{500} ($\gtrsim 0.68 r_{500}$) – 'Tier 1' and 'Tier 2', respectively. There are 23 groups for which this was possible, 15 of which are covered by the SDSS.

To supplement this sample, and to usefully extend the mass range for calibration of mass proxies, we add four groups and eight clusters with masses determined by Sanderson & Ponman. The cluster analysis is described in Sanderson & Ponman (2010, hereafter SP10). The full sample included 20 high flux *Chandra* clusters drawn from the flux-limited sample of Ikebe et al. (2002). The four groups were analysed in exactly the same way. Our superset sample therefore consists of 27 groups and clusters. However, in what follows we will exclude five groups for which galaxy membership was not well-defined (Section 2.2) or for which SDSS spectroscopic completeness was inadequate (Section 2.2.3), leaving a final sample of 22 systems.

The initial position and redshift of each of our groups was extracted from the NASA Extragalactic Database (NED).[2] We note that the cosmology used by S09 ($\Omega_{\rm m} = 0.24$, $\Omega_\Lambda = 0.76$ and $H_0 = 73\,{\rm km\,s^{-1}\,Mpc^{-1}}$) differs from that used here and in SP10. Given the low redshift of our groups ($z \lesssim 0.1$) the effect of $\Omega_{\rm m}$ and Ω_Λ is negligible. However, we apply a correction for H_0, scaling masses from S09 to $\mathrm{M}_\odot\, h_{70}^{-1}$.

[2] The NASA/IPAC Extragalactic Database (NED) is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

2.2 Optical

Optical counterparts were selected from the Seventh Data Release (DR7) of the SDSS (Abazajian et al. 2009). Specifically, we use data from the DR7 release of the New York University Value-Added Galaxy Catalogue (NYU-VAGC; Blanton et al. 2005). This catalogue features improvements on the original survey, including photometric calibration (Padmanabhan et al. 2008) and source identification. The NYU-VAGC also provides a set of *K*-corrected absolute magnitudes (Blanton & Roweis 2007) such that

$$M = m - 5\log\left(\frac{D_{\rm L}(z)}{10\,{\rm pc}}\right) - K(z) + 5\log h, \quad (1)$$

where $D_{\rm L}(z)$ is the luminosity distance to the galaxy, M and m are the absolute and apparent magnitude, respectively, and $K(z)$ is the *K*-correction. We use *K*-corrected catalogues where the *K*-correction has been found for filters shifted to $z = 0.1$.

The original DR7 is nominally spectroscopically complete to an apparent Petrosian magnitude limit of $m_r \approx 17.77$; however, this can vary somewhat across the sky. Following Berlind et al. (2006), we adopt a more conservative *r*-band magnitude limit of 17.5 to ensure more uniform coverage with an average completeness of ∼95 per cent as discussed below (Section 2.2.3). We further subset our galaxy sample by selecting only objects that meet the Blanton et al. galaxy criteria, i.e. have the bitmask `'VAGC_SELECT'>=4`, reducing potential contamination from spurious sources.

2.2.1 Group membership

To explore the impact of galaxy selection methods on our mass proxies, we construct two optical group samples. A 'volumetric' sample, based upon selecting galaxies within a quasi-cylindrical volume about each X-ray group, and a 'Friends-of-Friends' (FoF) sample constructed by performing an FoF analysis on the galaxies in the vicinity of each group and matching the extracted groups to the X-ray groups. This allows us to examine effects such as the imposition of a fixed metric aperture and the differing treatment of interlopers.

The volumetric sample is initially built by selecting galaxies from the NYU-VAGC spectroscopic sample within cones of radius 5 Mpc about the group positions drawn from NED. An initial velocity cut of ± 1000 km s^{-1} about the group redshift, also drawn from NED, is also applied.

Using galaxies from a smaller, 1 Mpc, aperture we then determine a velocity dispersion, σ, for the system using the *gapper* estimator from Beers et al. (1990). This estimator is unbiased in low-multiplicity systems and is robust against outliers. For a system of N galaxies each with a velocity v_i, we first order these velocities and determine the gap between pairs as $g_i = v_{i+1} - v_i$ for $i = 1, \ldots, N-1$. Each gap is weighted by its position within the ordered list, $w_i = i(N-i)$. The *gapper* estimator is then defined as

$$\sigma_{\rm gap} = \frac{\sqrt{\pi}}{N(N-1)} \sum_{i=1}^{N-1} w_i g_i. \quad (2)$$

As discussed by E04, if we assume that the brightest group galaxy (BGG) is stationary within the group halo, then $\sigma_{\rm gap}$ needs to be modified by a factor $\sqrt{N/(N-1)}$ giving

$$\sigma = \sqrt{\frac{N}{N-1}\sigma_{\rm gap}^2}. \quad (3)$$

A refined volumetric group membership is then constructed iteratively by applying this 3σ clip until the number of galaxies within 1 Mpc converges (i.e. velocity dispersion converges). The galaxy sample within the 5 Mpc cone is then refined by applying a 3σ clip using the same value of σ.

We then determine the group centroid using these galaxies to provide a consistent centroid definition (Section 2.2.2) and repeat this selection using this updated group position, rather than the NED coordinates.

We assume that our galaxy sample consists of two populations, an interloper-contaminated group population within a 1 Mpc aperture, and a local background which we determine within an annulus of radius 3–5 Mpc about the group centre. Where possible, we use this local background to statistically subtract interloper contamination from our mass proxies. This annulus was chosen to be large enough to reduce any group contribution to this local background even for the largest systems (e.g. $r_{200} \sim 2$ Mpc for $M_{200} \sim 1 \times 10^{15}\,{\rm M}_\odot$, assuming a mean density $200\rho_c(z)$ at $z = 0.1$). This background may include infalling galaxies or field galaxies in the Hubble flow that have similar redshifts to the group itself. Mamon, Biviano & Murante (2010) analysed an ensemble of haloes from a cosmological simulation and found the surface density of interlopers within a σ clipped volume to be approximately uniform with radius, so our background subtraction should be effective so long as foreground or background structures contribute equally to our source and background regions. The application of this background correction is described in Section 3.

Due to sparse spectroscopic coverage for one group our calculation fails to converge on a velocity dispersion and group membership. We exclude this group from our analysis, reducing the sample to 26 groups.

To generate the FoF group sample, we follow the algorithm described by E04. This was originally developed for the 2-degree Field Galaxy Redshift Survey (Colless et al. 2001). We use the E04 linking length calibration with modifications for the appropriate SDSS luminosity function (LF) and magnitude limits. We apply this to all galaxies within a cone with a radius of at least 5 Mpc and extending to a redshift of 0.2 about each volumetric group. This volume will contain more than one FoF group for a given X-ray system. We select the central group by matching the NED position to the centre of identified groups within a range of ± 0.05 in redshift.

2.2.2 Group centres

We identify the centres of both volumetric and FoF groups using an iterative centre of light approach similar to that described in Robotham et al. (2011). This iterative approach determines a weighted centroid of the galaxy distribution. In addition to weighting each galaxy by its luminosity, we also weight each galaxy by its velocity offset from the mean. At each step, the galaxy furthest from this centroid is discarded. This process is repeated until only two galaxies remain, and the brightest of these is then taken as the group centre. With the volumetric sample this is performed only with galaxies within the innermost 1 Mpc, whilst all member galaxies are used for the FoF sample.

Group redshifts are the median redshift of galaxies within 1 Mpc, or the median redshift of all galaxies for the volumetric and FoF samples, respectively.

2.2.3 Spectroscopic completeness and survey coverage

Whilst the SDSS spectroscopy is nominally $\gtrsim$85 per cent complete to $m_r = 17.77$ (Blanton et al. 2005), variation in the target selection function, as well as incompleteness due to fibre collisions

and obscuration from bright foreground objects, can lead to a variable level of completeness across the sky (Berlind et al. 2006). Our initial apparent magnitude cut of $m_r \leq 17.5$ allows us to be confident of greater, and more uniform, completeness.

Correcting for fibre collisions is especially important in regions of high density such as the centres of galaxy groups and clusters. This problem is moderated slightly for the SDSS due some overlap in its tiling pattern allowing for repeat visits to regions of the sky. However, as this overlap covered only a small fraction of the observed field, fibre collisions still need to be considered. Failure to account for this would result in an underestimate of galaxy group richness and galaxy overdensities within these regions. We use the NYU-VAGC data sets in which each collided galaxy is assigned the spectroscopic redshift of the galaxy it collided with (Blanton et al. 2005). This correction, found by Blanton et al. (2005) to affect ∼6 per cent of galaxies globally, is used for 10.9 per cent of the galaxies in our sample within 1 Mpc of the group centre. As we focus on regions with known groups or clusters, i.e. high-density galaxy fields, this increase compared to the global average is to be expected.

The incidence of collided galaxies is larger in the dense central regions, and these central collisions are found in 12 of the final 22 groups, with, on average, two corrected galaxies in each. In case the 'cloning' of redshifts in these systems introduces any significant bias into their calculated velocity dispersions, we will pay special attention to them in the analysis presented in Section 5.3.

We assess the level of any remaining spectroscopic incompleteness by direct comparison of the spectroscopic and photometric data. Selecting both photometric and spectroscopic galaxies (within our apparent magnitude limited sample) within a cone about a point on the sky, we can define completeness at each position as the ratio of the number of galaxies with a redshift to the total number photometric galaxies available. To control the statistical uncertainty on each point, yet ensure we have resolution as high as possible in dense regions, the size of the cone is adjusted to contain at least 25 spectroscopic galaxies. In most cases, we find completeness close to unity, with an average value of ∼95 per cent within our groups. The inverse of the local completeness is included as a weight on each galaxy when applied to the majority of our methods as described in Section 3. As a dynamical mass estimator is robust to the effects of incompleteness (assuming no velocity substructures), we do not attempt to correct for incompleteness in the case of dynamical mass proxies.

We also use the photometric data to derive a correction for fields located on the edge of the survey. Using a coarsely smoothed map of the photometric galaxy density we define the survey edge to be a contour $2\sigma_\Sigma$ below the mean surface density, where the mean and standard deviation, σ_Σ, are found iteratively, excluding regions of high (the group) and low (beyond the survey edge) density. Fig. 1 illustrates an example completeness map for one of our groups which falls close to the SDSS survey boundary. We use this knowledge of the survey boundary to account for any missing area within each group, as described in Section 3.

To ensure that we do not use data requiring very large corrections for spectroscopic completeness or survey coverage, we examine the 26 groups remaining in the current sample, and exclude those which have a central or background completenesses <70 per cent (within a 1 Mpc aperture and a 3–5 Mpc annulus, respectively), or have <70 per cent and <50 per cent footprint coverage, respectively, in the source and background regions. This excludes an additional four groups, reducing the sample to 23 systems, four of which intersect the survey edge. We exclude one last group due to a heavily

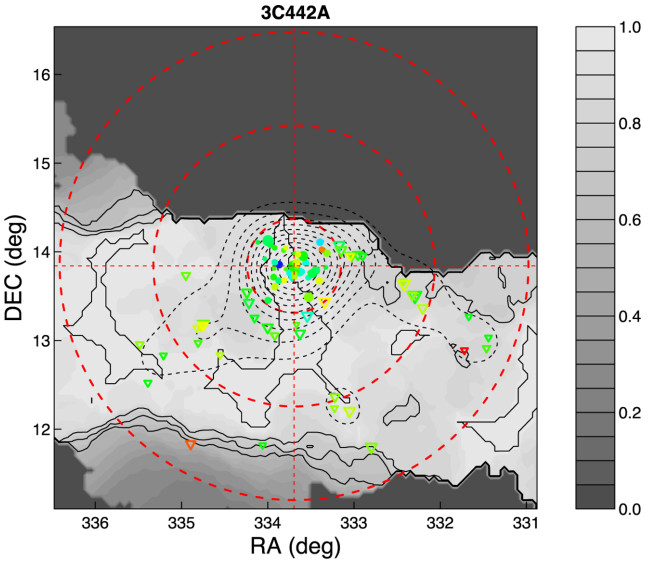

Figure 1. Completeness map for the group 3C 442A. Grey-scale contours indicate the measured spectroscopic completeness as described in the text, with black contours delineating 100, 80, 70 and 60 per cent completeness levels. Dashed red circles indicate our 1, 3 and 5 Mpc boundaries representing the areas within which membership and background are determined, respectively. The points represent the galaxies that pass through the 3σ clip, filled circles represent member galaxies, open triangles all galaxies beyond 1 Mpc, whilst colour represents redshift (galaxies at larger redshift are redder in the range $0.022 \lesssim z \lesssim 0.030$).

contaminating background structure that could not be satisfactorily excluded. The final group sample consists of 22 groups. Their X-ray and initial optical properties are described in Table 1.

2.2.4 *Luminosity completeness*

As our optical data are drawn from an apparent magnitude-limited survey, which results in a redshift-dependent absolute magnitude limit, a final correction factor is needed to rescale results to the same absolute magnitude for those mass proxies which scale with the number or luminosity of galaxies.

To correct for this, we integrate over the galaxy LF to calculate two factors,

$$c_{\rm N} = \frac{\int_{L_r^{\rm faint}(z)}^{L_r^{\rm bright}} \Phi(L_r){\rm d}L_r}{\int_{L_r^{\rm faint}}^{L_r^{\rm bright}} \Phi(L_r){\rm d}L_r}, \tag{4}$$

and

$$c_{\rm L} = \frac{\int_{L_r^{\rm faint}(z)}^{L_r^{\rm bright}} L_r\Phi(L_r){\rm d}L_r}{\int_{L_r^{\rm faint}}^{L_r^{\rm bright}} L_r\Phi(L_r){\rm d}L_r}, \tag{5}$$

to extrapolate number ($c_{\rm N}$) and luminosity ($c_{\rm L}$) estimates to a fixed absolute magnitude. We use $\Phi(L)$ to be a standard Schechter LF (Schechter 1976) with parameters from the *r*-band cluster LF of Popesso et al. (2005) ($M_r^* - 5\log h = -21.35 \pm 0.19$ and $\alpha = -1.30 \pm 0.06$). We set the faint limit to be $M_r = -16.5$, comparable to the faintest absolute magnitude of galaxies within our

Table 1. Our group + cluster sample.

Group	z^a	RA^a (deg)	Dec^a (deg)	$N_{1\mathrm{Mpc}}{}^b$	σ^c (km s^{-1})	T_X (keV)	$M_{500,\mathrm{X}}$ (10^{13} M$_\odot$)	Reference^d
3C 442A	0.0261	333.70	13.841	48	435^{+47}_{-33}	1.34 ± 0.04	$4.07^{+0.23}_{-0.42}$	(1)
Abell 0160	0.0428	18.32	15.516	50	818^{+97}_{-68}	1.68 ± 0.10	$8.24^{+1.11}_{-1.15}$	(1)
Abell 1177	0.0322	167.44	21.759	28	311^{+50}_{-29}	$1.37^{+0.06}_{-0.07}$	$5.51^{+0.88}_{-0.76}$	(1)
Abell 1275	0.0628	172.50	36.559	21	436^{+83}_{-40}	$1.46^{+0.08}_{-0.07}$	$7.20^{+3.13}_{-1.74}$	(1)
Abell 1367	0.0214	176.01	19.950	170	833^{+54}_{-43}	3.22 ± 0.18	13.20 ± 2.88	(2)
Abell 1692	0.0836	198.06	−0.974	26	655^{+115}_{-64}	$2.61^{+0.16}_{-0.24}$	$10.12^{+3.12}_{-1.99}$	(1)
Abell 1795	0.0629	207.22	26.593	62	742^{+64}_{-42}	$5.62^{+0.36}_{-0.35}$	53.00 ± 7.50	(2)
Abell 1991	0.0588	223.65	18.688	42	611^{+82}_{-48}	$2.68^{+0.10}_{-0.08}$	$13.97^{+2.61}_{-1.98}$	(1)
Abell 2029	0.0774	227.75	5.783	72	1450^{+119}_{-75}	8.96 ± 0.30	71.50 ± 17.10	(2)
Abell 2092	0.0668	233.31	31.145	34	567^{+85}_{-46}	$1.67^{+0.13}_{-0.12}$	$9.33^{+1.89}_{-1.69}$	(1)
Abell 2142	0.0903	239.59	27.334	58	1340^{+170}_{-119}	$9.50^{+0.43}_{-0.42}$	125.00 ± 14.60	(2)
Abell 2147	0.0368	240.55	15.918	134	944^{+68}_{-52}	3.69 ± 0.18	26.70 ± 3.72	(2)
Abell 2199	0.0304	247.16	39.553	150	841^{+49}_{-33}	$4.50^{+0.20}_{-0.24}$	31.50 ± 4.25	(2)
Abell 85	0.0564	10.46	−9.303	73	872^{+81}_{-58}	6.64 ± 0.20	47.40 ± 5.05	(2)
MKW04	0.0200	181.11	1.896	70	518^{+58}_{-45}	1.58 ± 0.09	$5.06^{+0.74}_{-0.71}$	(1)
NGC 4104	0.0277	181.66	28.174	50	461^{+43}_{-28}	$1.41^{+0.09}_{-0.06}$	$5.06^{+0.57}_{-0.55}$	(1)
NGC 4325GROUP	0.0255	185.78	10.621	29	305^{+47}_{-25}	1.00 ± 0.02	3.65 ± 0.44	(2)
NGC 5098	0.0365	200.05	33.324	48	523^{+55}_{-33}	0.96 ± 0.04	$2.09^{+0.29}_{-0.48}$	(1)
NGC 6338GROUP	0.0289	258.85	57.411	67	628^{+63}_{-43}	$2.03^{+0.12}_{-0.11}$	8.99 ± 0.51	(2)
RXJ1022.1+3830	0.0546	155.61	38.539	27	649^{+96}_{-57}	$1.94^{+0.20}_{-0.14}$	$8.34^{+1.37}_{-1.46}$	(1)
RXJ1159.8+5531	0.0806	179.97	55.535	14	365^{+68}_{-25}	$1.84^{+0.14}_{-0.08}$	$8.66^{+3.23}_{-1.17}$	(1)
UGC05088	0.0274	143.50	33.991	14	234^{+54}_{-18}	0.81 ± 0.03	$1.54^{+0.38}_{-0.25}$	(1)

Notes. [a]Coordinates for the volumetric sample estimated as described in Section 2.2.2.
[b]Group multiplicity within 1 Mpc and brighter than $m_r = 17.5$.
[c]Velocity dispersions calculated for the innermost 1 Mpc galaxies of the volumetric selection.
[d]Masses and temperatures drawn from (1) Sun et al. (2009, S09); (2) Sanderson & Ponman (2010, SP10).

sample and converted to luminosity assuming $M_{r,\odot} = 4.67$.[3] The correction is therefore ∼1 for the closest groups, and reaches values of $c_\mathrm{N} = 0.11$, $c_\mathrm{L} = 0.65$ for our most distant, at $z \approx 0.09$.

3 MASS PROXIES

3.1 Richness and luminosity

One of the simplest mass proxies available to us is the number of galaxies brighter than a specific luminosity contained within a galaxy group. Through hierarchical formation we expect larger dark matter haloes to have formed through the assimilation of smaller structures, and the total number of galaxies will be conserved during this process. Whilst there are processes that can reduce the final richness of groups, such as the orbital decay and merger of large galaxies, the impact of this on richness would be limited by the large dwarf populations in groups. In the absence of major trends in star formation efficiency (SFE) and trends in the LF parameters with halo mass, we would naively expect richness to scale approximately linearly with mass. In practice, this is not the case; SFE, and therefore stellar mass fraction, have been shown to vary with halo mass (e.g. Moster et al. 2010; Leauthaud et al. 2012), whilst LFs have been found to become significantly more 'field'-like in lower mass haloes (e.g. Hansen et al. 2005; Robotham, Phillipps & de Propris 2010; Zandivarez & Martínez 2011), both of which will affect the number of galaxies observed above a given luminosity. We discuss the impact these have on our naive expectations in Section 5.1.

Richness-based mass proxies have previously been studied by a number of different authors. Budzynski et al. (2012), for example, use a mass–richness relation to determine the masses of their high mass group and cluster sample ($M_{500} > 10^{13.7}$ M$_\odot$). Other studies examine the mass–richness relation for a subset of galaxy types, such as red sequence galaxies (Rozo et al. 2009), and sources of scatter therein (Rozo et al. 2011), as well as Andreon & Hurn (2010) who similarly select only red galaxies. Rykoff et al. (2012) extend the richness estimator of Rozo et al. (2009) to include blue galaxies, finding that it increases the observed scatter in the L_X-richness relation. However, rather than restricting our analysis to a single class of galaxy, which would limit the diversity of groups suitable for analysis, we use all galaxies within each group.

To avoid the circularity involved in counting galaxies within some radius (e.g. an overdensity radius) which scales with mass, we employ a simple aperture-limited richness for our volumetric sample. Richness is defined as the number of galaxies within 1 Mpc of the group centre, corrected for incompleteness. Each galaxy's contribution within the aperture is weighted by the inverse of its local spectroscopic completeness, and the number counts, after background subtraction, are corrected to our standard absolute magnitude range as described in Section 2.2.4 above. We make no correction for projection effects or for the imposition of a fixed aperture. Our aim here is to keep the procedure simple and to calibrate out these effects by comparison with the X-ray masses.

[3] As listed by C. N. A. Willmer, http://mips.as.arizona.edu/~cnaw/sun.html

To correct for interloper contributions we first estimate the surface number density in a background annulus of radius 3–5 Mpc. Again, each galaxy contribution is weighted by its local spectroscopic completeness. If the annulus intersects the survey edge, we estimate the area that is covered and determine the surface density within this area. Using this surface density, we estimate the number of background galaxies within the 1 Mpc aperture (again, correcting for the survey footprint if necessary). We subtract this background estimate from our total richness and rescale to a fixed absolute magnitude limit using equation (4) to determine our final corrected volumetric richness estimate, $\tilde{N}_{1\,\mathrm{Mpc}}$.

For our FoF sample, the procedure is simpler. Here no background subtraction is performed – it is assumed that background contamination is negligible – and the FoF multiplicity (N_{FoF}) is corrected for spectroscopic and survey incompleteness and scaled to our standard absolute magnitude, to give the corrected richness estimate, $\tilde{N}_{\mathrm{FoF}}$.

A closely related mass proxy to richness is the total optical luminosity of a group. This has the advantage over richness of being less sensitive to variations in the faint-end slope of the LF. Furthermore, galaxy mergers conserve luminosity but not number, although in practice it is known that mergers and tidal interactions can remove stars from galaxies, forming an intracluster light (ICL; see e.g. Gonzalez, Zabludoff & Zaritsky 2005; Zibetti et al. 2005) component. Our SDSS-derived luminosities do not include any contribution from ICL, and rather than attempting to correct for any trends in ICL, we assume that they can be calibrated out in our mass–luminosity analysis below. We convert absolute magnitudes to luminosities in solar units assuming $\mathrm{M}_{r,\odot} = 4.67$, and treat background and correction as for richness. The final luminosity-based mass proxies, $L_{1\,\mathrm{Mpc}}$ and L_{FoF}, are then derived by rescaling to our standard absolute magnitude using the c_{L} factor from equation (5).

3.2 Overdensity

It is has been established that dark matter haloes in simulations are generally well represented by Navarro, Frenk and White density (NFW; Navarro, Frenk & White 1996) profiles,

$$\rho(r) = \frac{\rho_0}{\frac{r}{r_{\mathrm{s}}}\left(1+\frac{r}{r_{\mathrm{s}}}\right)^2}, \tag{6}$$

where r_{s} is the scale radius and ρ_0 is the normalization, or characteristic density of the NFW profile.

Under the assumption that galaxies trace mass, it is possible to use the galaxy density profile to infer the total mass of the system. A similar approach was employed by Hansen et al. (2005), who determined galaxy surface densities by counting galaxies within cylinders of increasing radius about each group or cluster. By applying a scale factor, these surface densities were converted into 3D densities and, comparing to the field density based on the Blanton et al. (2003) SDSS LFs, characteristic overdensity radii were derived. These radii, $r_{200}^{\mathcal{N}}$, were defined to be the radius at which the estimated galaxy density was $200/\Omega_{\mathrm{m}}$ times the mean galaxy density, where the factor of Ω_{m} allows conversion from mean to critical density.

We expand on this approach by fitting a projected NFW profile to the observed galaxy number distribution and using the resultant NFW profile to determine the radius at which the density of galaxies is $500/\Omega_{\mathrm{m}}$ times that derived from the global LF. This radius should be equivalent to the radius enclosing a mean density of 500 times the critical density. Using a maximum likelihood method, we fit the projected NFW profile derived by Bartelmann (1996),

$$\Sigma_{\mathrm{NFW}}(x) = \frac{2\rho_0 r_{\mathrm{s}}}{x^2-1} f(x), \tag{7}$$

where $x = r/r_{\mathrm{s}}$ and

$$f(x) = \begin{cases} 1 - \frac{2}{\sqrt{x^2-1}} \arctan\sqrt{\frac{x-1}{x+1}} & x > 1 \\ 1 - \frac{2}{\sqrt{1-x^2}} \operatorname{arctanh}\sqrt{\frac{1-x}{1+x}} & x < 1 \\ 0 & x = 1 \end{cases}.$$

For the volumetric sample, we add to this fit a background term, Σ_{bg}, to account for any background in the field, i.e.

$$\Sigma(r) = \Sigma_{\mathrm{NFW}}(r) + \Sigma_{\mathrm{bg}}. \tag{8}$$

Incorporating this background term, we fit the galaxy distribution for each group within its full 5 Mpc, 3σ cone. For the FoF sample, the profile is fit to the FoF-linked galaxies assuming no background contribution.

The fit to the observed number distribution of galaxies involves using equation (8) to predict the number of galaxies within each annulus of width dr, centred at projected radius r,

$$\mathrm{d}N(r) = 2\pi A(r) r \Sigma(r) \mathrm{d}r. \tag{9}$$

The constant 2π assumes an annulus of radius r and width dr is fully covered, and in these cases $A(r) = 1$. As already discussed, there are instances where the group intersects the survey edge. For these groups, the factor $A(r) \leq 1$ is used to rescale the area of the annulus as required.

The likelihood, $\mathcal{L}_i$, of observing a galaxy i at radius r_i is

$$\mathcal{L}_i = \frac{2\pi A(r_i)\Sigma(r_i)}{\int_0^{r_{\mathrm{max}}} 2\pi A(r)\Sigma(r)\mathrm{d}r}, \tag{10}$$

where r_{max} is the largest projected radius of the galaxies included in the fit. The likelihood function that we maximize with respect to r_{s} and Σ_{bg} is then

$$\mathcal{L} = \prod_i^N \mathcal{L}_i. \tag{11}$$

The full analysis is an iterative process, in which we initially use all galaxies within a group's volume, with each galaxy weighted by the inverse of its local completeness. The fitted projected NFW and background terms are then renormalized to reproduce the observed number of galaxies used in the fit.

The group mass is then inferred by integrating the 3D NFW density profile, equation (6), using the values of r_{s} and ρ_0 determined from the fit. This gives a mean galaxy density profile from which we estimate r_{500} as the radius within which the mean number density is $500/\Omega_{\mathrm{m}}$ times the mean galaxy number density. We determine the mean number density using a global LF for the SDSS with r-band parameters of $M_r^* - 5\log h = -20.44 \pm 0.01$, $\alpha = -1.05 \pm 0.01$ and $\phi^* = (1.49 \pm 0.04) \times 10^{-2} h^3\ \mathrm{Mpc}^{-3}$ from Blanton et al. (2003), and integrate from $M_r = -25$ to an absolute magnitude determined by the apparent magnitude limit at the group's redshift.

Using this initial estimate of r_{500} and the fit results as new initial guesses for r_{s} and Σ_{bg}, we then repeat the fit excluding the central galaxies ($r < 0.1 r_{500}$). This radial cut excludes central galaxies, which may be affected by orbital decay and galaxy merging at the centre of the group potential, and may therefore depart from the expected NFW distribution. We run this core-excised fit twice, iterating r_{500}.

We adopt a similar approach to define a *luminosity overdensity*. As with the richness/luminosity estimators, this extension provides an estimator which should be more robust against variations in the faint-end slope of the LF, since this contains many galaxies but not much light. The above analysis is simply modified to weight each galaxy in the likelihood fit by its luminosity. When determining luminosity overdensity we also explicitly add the BGG luminosity, which can represent a substantial fraction of the group total for poor groups, and estimate the mean global luminosity density from the Blanton et al. (2003) LF.

We note that this approach assumes that an overdensity in the baryonic properties (e.g. galaxy number) of a group relates directly to the same overdensity in mass, essentially assuming that light-traces-mass. As demonstrated by the need to excise the core, this is not necessarily the case. Indeed, it has been shown that galaxy density profiles are typically a factor of 2 less concentrated than those of dark matter profiles (e.g. Budzynski et al. 2012). As we normalize galaxy profiles to recover the observed number and total light of galaxies, this concentration difference will trade against normalization, ultimately reducing the observed group overdensity radius compared to what would be expected if the galaxies were concentrated as the underlying dark matter. This should however be a systematic bias we can calibrate out. More importantly, effects such as mass-dependent variation in SFE will likely introduce substantial bias into this analysis which we will discuss more in Section 5.2.

3.3 Dynamical mass

Dynamical mass estimation is one of the oldest and most well established techniques. The Virial Theorem, applied to a stable system, leads directly to

$$M \propto r\sigma^2. \tag{12}$$

Eke et al. (2004), calibrating this relation on simulated clusters for the 2dF Galaxy Redshift Survey, with a cosmology of $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ with $h = 1$, find

$$M_{\rm vir} = 5\frac{r_{\rm rms}\sigma^2}{G}\quad , \tag{13}$$

where $r_{\rm rms}$ is the rms group radius and σ is the velocity dispersion found using the *gapper* estimator described in Section 2.2.1. For the same H_0 with a different set of simulations, Robotham et al. (2011) find a substantially higher constant with

$$M_{\rm halo} = 10\frac{r_{50}\sigma^2}{G}, \tag{14}$$

where r_{50} is the radius containing 50 per cent of a group's galaxies. This constant is likely due to the different definition of radius.

Alternatively, this can be cast completely in terms of the velocity dispersion, σ, by assuming $r_\Delta \propto \sigma$ (e.g. Carlberg et al. 1997; Ramella et al. 2004) such that

$$M_\Delta = 3\frac{r_\Delta\sigma^2}{G} = 3\sqrt{\frac{6}{\Delta}}\frac{\sigma^3}{H(z)G}, \tag{15}$$

where Δ is the overdensity, relative to the critical density, enclosed within r_Δ.

Despite the fact that these methods are well established, we include them to investigate the possible biases that have been reported in dynamical mass estimates of poor groups, as discussed in Section 1. For each of these estimators, we reduce the proxies to $M_{500} \propto r_{\rm rms}\sigma^2$ and $M_{500} \propto \sigma^3$, respectively, taking radii and velocity dispersion within 1 Mpc for the volumetric sample, and using all member radii and velocities for the FoF sample.

3.4 Radii

At constant density, it is easy to see that the mass and radius of groups are related by $M \propto r^3$. We initially examined a number of different estimates for the projected group radius, such as the half-light radius and harmonic mean radius, finding little difference in their behaviour. We use the root-mean-square radius, $r_{\rm rms}$, of the group galaxies.

For the FoF sample, $r_{\rm rms}$ is simply the rms radius of all the linked group galaxies. For the volumetric sample, we use all velocity-selected galaxies within the 1 Mpc cone. Each galaxy is again weighted by its local completeness. This estimator ignores the expected background contamination, since we have no way of knowing exactly which galaxies are interlopers. Since the distribution of interlopers should be fairly uniform, we expect that this will bias estimated radii to larger values. When applied to a volumetric sample, both this bias and the use of an aperture turn out to be serious flaws for this mass proxy.

4 STATISTICAL ANALYSIS

Our aim is to calibrate the relationship between system mass and each of our mass proxies, and to examine the statistical performance of each method. In each case, we evaluate the strength of correlations using the Spearman rank correlation coefficient, and fit the mass–proxy relations with power laws of the form

$$\log_{10}\left(\frac{M_{500}}{10^{14}\,\mathrm{M}_\odot}\right) = \alpha\log_{10}\left(\frac{x}{x_0}\right) + \beta, \tag{16}$$

where M_{500} is the X-ray-determined mass and x is the relevant observable. To reduce correlation between the fitted slope and intercept we pivot the fit about $10^{14}\,\mathrm{M}_\odot$ and $x_0 = \bar{x}$.

4.1 Regression methods

To avoid arriving at biased estimates of calibrated relations, it is important to use the most appropriate regression method. Different methods make different assumptions regarding the dependent and independent variables and optimize scatter differently. These differences may result in biased estimates of any relation.

To explore this, we examine the performance of two different techniques which have been widely used – the frequentist BCES regression methods (Akritas & Bershady 1996) and the Bayesian fitting approach of Kelly (2007) – and compare them with a generative modelling technique of our own.

The BCES estimators are a general extension of the ordinary least squares estimator capable of accounting for intrinsic scatter and measurement errors in both axes. A number of forms of BCES are available: (Y|X) and (X|Y) regression which distinguish dependent and independent variables, and symmetric bisector and orthogonal techniques.

The Kelly (2007) estimator is a Bayesian linear regression estimator that models the mass–proxy relation, stressing the importance of correctly handling statistical errors. The method not only determines the optimum regression line but also provides an estimate of the intrinsic scatter in the relation.

The generative fitting method attempts to model the statistics of the process that produces measured mass and proxy values. The procedure is similar to some of the methods discussed in Hogg,

Table 2. Results of regression tests – recovered slopes and their standard errors for each regression method.

Method	Mean recovered slope
Input	0.8
BCES(Y|X)	0.686 ± 0.005
BCES(X|Y)	0.808 ± 0.007
BCES(Bi)	0.745 ± 0.006
BCES(Orth)	0.727 ± 0.006
Kelly	0.687 ± 0.005
Generative	0.730 ± 0.008

Bovy & Lang (2010); it is also similar in spirit, if not in detail, to the method of Kelly (2007). However, whilst Kelly (2007) assumes an 'independent' variable upon which the 'dependent' second variable is conditioned, our generative model treats the two variables symmetrically. Further details of the generative fitting method can be found in Appendix A.

To decide which of these methods is the best method to use, we generate an ensemble of 200 mock mass–proxy data sets using a generative model and, through application of each fitting method, evaluate their accuracy in recovering the input relation. Our aim is to use a regression with the least biased slope, allowing, on average, accurate masses be recovered from a mass–proxy relation. Alternatively, we could optimize to a relation that recovers the lowest intrinsic scatter. Due to the sample size, we expect statistical errors and errors on the calibration to contribute significantly to the scatter, therefore we prioritize recovering an unbiased relation over minimizing the intrinsic scatter about that relation.

Each mock data set consists of 22 mass values drawn from a log-uniform distribution with $13 \lesssim \log_{10}(M/\mathrm{M}_\odot) \lesssim 15$. We assume a mass–proxy relation as per equation (16) with slope $\alpha = 0.8$ and intercept $\beta = 5$. From this, proxy values are found as

$$\log_{10}(x) = 1/\alpha \log_{10}(M) - \beta/\alpha + N(0, \sigma_{\rm int}),$$

where $N(0, \sigma_{\rm int})$ models intrinsic scatter in the proxy about the relation, assuming proxies are scattered as a Gaussian of width $\sigma_{\rm int} = 0.3$ dex. The final step of constructing the mock data set is to add statistical errors. To do this, we rank both mock and observed mass–proxy pairs by mass and match them. The mock data are then assigned errors in mass and proxy based on the corresponding errors in the observational sample, where the proxy errors are drawn from the volumetric σ^3 proxy.

Each of the regression techniques are applied to the mass–proxy data for all 200 data sets. We collect the estimated slopes from each ensemble and determine the mean slope and standard error on this mean for each regression method. The results in Table 2 show that all methods apart from the frequentist BCES(X|Y) estimator return slopes which exhibit significant bias relative to the input value of 0.8.

Given these results, we adopt the BCES(X|Y) estimator to fit the mass–proxy calibration relations used for the remainder of this study.

4.2 Errors and scatter analysis

Measurement errors on masses are taken from the studies of S09 and SP10. Statistical errors on the proxies are derived in one of two ways. For the richness proxy, we simply assume Poisson noise on galaxy number counts. All other values have errors defined by a 68 per cent confidence interval from a bootstrap analysis. For each group, and without redefining the galaxy selection (i.e. we do not repeat the FoF analysis), we resample from its input galaxy catalogue a new, random set of galaxies, with replacement, of equal size to the original. We then repeat our analysis to determine a revised set of mass proxies. This is performed 1000 times for each group, allowing a distribution of proxy values to be found.

We report all errors and scatter estimates as fractional errors in dex, where the statistical errors on each group are converted to dex[4] using $\sigma_{\log x} = \sigma_x/\ln(10)x$. Errors on our fitted calibration parameters are derived by bootstrap resampling from the group mass–proxy pairs. The resulting fits are presented in Table 3.

We illustrate the error on our fits as the shaded regions on Fig. 2. At each point along the proxy axis, we calculate a mass distribution using the spread of calibration parameters from the bootstrap resampling. The error region is then a 68 per cent (1σ) interval bounding the 16th to 84th percentiles in mass about the regression line.

One aim of this work is to explore the statistical performance of these mass proxies. We approach this by attempting to quantify the minimum error one would see if applying these relations to perfect data. That is, assuming no measurement error on a given proxy, how much uncertainty in a mass estimate would be introduced by the intrinsic scatter about these relations and the uncertainty in our calibration?

We estimate intrinsic scatter as

$$\sigma^2_{\log M,\rm int} = \sigma^2_{\log M,\rm tot} - \sigma^2_{\log M,\rm stat} - \alpha^2\sigma^2_{\log x,\rm stat}, \quad (17)$$

where $\sigma_{\log M,\rm tot}$ is the total fractional scatter in mass observed about the relation (i.e. the rms fractional residuals) and $\sigma_{\log M,\rm stat}$ the mean fractional statistical error in the mass (0.08 dex). The term $\alpha^2\sigma^2{}_{\log x,\rm stat}$ accounts for the additional scatter in mass about the relation introduced by the statistical scatter of the proxy and propagated into the intrinsic scatter as $\sigma^2_{\log M} = (\mathrm{d}\log_{10}(M)/\mathrm{d}\log_{10}(x))^2\sigma^2_{\log x}$.

The mean uncertainty in mass introduced by the calibration is found by

$$\sigma^2_{\log M,\rm cal} = \frac{1}{N}\sum_i^N \left[\left(\log_{10}\left(\frac{x_i}{x_0}\right)\sigma_\alpha\right)^2 + \sigma^2_\beta + 2\log_{10}\left(\frac{x_i}{x_0}\right)\mathrm{Cov}(\alpha,\beta)\right], \quad (18)$$

where $\mathrm{Cov}(\alpha, \beta)$ is the covariance of the relation parameters, accounting for any correlation in the errors. The sum $\sigma^2_{\log M,\rm sys} = \sigma^2_{\log M,\rm int} + \sigma^2_{\log M,\rm cal}$ gives the minimum fractional variance in mass one would expect for a given proxy. Our scatter analysis is also summarized in Table 3.

5 RESULTS AND DISCUSSION

In this section, we present the main results of this investigation. Using the statistical techniques described in the previous section, we calibrate each mass–proxy relation and examine its statistical performance. Where possible, we also make comparisons with theoretical expectations to try to better understand these relations.

5.1 Richness and luminosity

The top row of Fig. 2 shows the results of the richness and luminosity proxies, with fit estimates in Table 3. We find that for both galaxy

[4] For example, 0.1 dex is a factor of $10^{0.1}$.

Table 3. BCES regression results for our sample corresponding to the best-fitting regression lines in Fig. 2.

Method	α	β	$\log_{10}(x_0)$	ρ^a	$\sigma_{\log x,\,\mathrm{stat}}{}^b$	$\sigma_{\log M,\,\mathrm{tot}}{}^b$	$\sigma_{\log M,\,\mathrm{int}}{}^{b,c}$	$\sigma_{\log M,\,\mathrm{cal}}{}^b$	$\sigma_{\log M,\,\mathrm{sys}}{}^{b,d}$
				Volumetric selection					
Richness	$1.61^{+0.21}_{-0.23}$	$0.21^{+0.05}_{-0.04}$	2.24	0.93	$0.08^{+0.01}_{-0.01}$	0.22	$0.17^{+0.09}_{-0.04}$	$0.08^{+0.02}_{-0.01}$	$0.19^{+0.08}_{-0.04}$
Luminosity	$1.73^{+0.24}_{-0.26}$	$0.22^{+0.06}_{-0.05}$	11.94	0.88	$0.11^{+0.01}_{-0.01}$	0.30	$0.21^{+0.13}_{-0.05}$	$0.10^{+0.03}_{-0.02}$	$0.23^{+0.12}_{-0.05}$
Overdensity (N)	$1.25^{+0.10}_{-0.12}$	$0.29^{+0.05}_{-0.05}$	14.32	0.90	$0.17^{+0.02}_{-0.02}$	0.27	$0.13^{+0.12}_{-0.06}$	$0.07^{+0.01}_{-0.01}$	$0.15^{+0.09}_{-0.06}$
Overdensity (L)	$1.14^{+0.11}_{-0.12}$	$0.31^{+0.06}_{-0.05}$	14.34	0.89	$0.19^{+0.02}_{-0.02}$	0.26	$0.12^{+0.11}_{-0.05}$	$0.07^{+0.02}_{-0.01}$	$0.14^{+0.08}_{-0.05}$
Dynamical ($r_{\mathrm{rms}}\sigma^2$)	$1.26^{+0.18}_{-0.23}$	$0.30^{+0.06}_{-0.06}$	8.50	0.84	$0.09^{+0.01}_{-0.005}$	0.30	$0.26^{+0.08}_{-0.06}$	$0.10^{+0.03}_{-0.02}$	$0.28^{+0.09}_{-0.06}$
Dynamical (σ^3)	$0.94^{+0.13}_{-0.17}$	$0.39^{+0.07}_{-0.07}$	8.70	0.85	$0.13^{+0.01}_{-0.01}$	0.30	$0.26^{+0.08}_{-0.05}$	$0.10^{+0.03}_{-0.02}$	$0.28^{+0.09}_{-0.05}$
Radius	$12.60^{+12.33}_{-15.38}$	$0.09^{+0.23}_{-0.19}$	2.75	0.66	$0.03^{+0.004}_{-0.004}$	0.54	$0.31^{+0.33}_{-0.26}$	$0.21^{+1.91}_{-2.12}$	$0.37^{+1.95}_{-2.11}$
				FoF selection					
Richness	$1.03^{+0.13}_{-0.13}$	$0.34^{+0.07}_{-0.06}$	2.63	0.93	$0.06^{+0.01}_{-0.01}$	0.24	$0.21^{+0.07}_{-0.04}$	$0.08^{+0.02}_{-0.01}$	$0.23^{+0.07}_{-0.04}$
Luminosity	$1.06^{+0.15}_{-0.15}$	$0.33^{+0.08}_{-0.06}$	12.33	0.91	$0.08^{+0.01}_{-0.01}$	0.26	$0.23^{+0.08}_{-0.04}$	$0.09^{+0.02}_{-0.02}$	$0.25^{+0.08}_{-0.04}$
Overdensity (N)	$1.24^{+0.11}_{-0.12}$	$0.30^{+0.07}_{-0.06}$	14.33	0.88	$0.15^{+0.02}_{-0.01}$	0.29	$0.22^{+0.11}_{-0.05}$	$0.08^{+0.02}_{-0.01}$	$0.23^{+0.11}_{-0.05}$
Overdensity (L)	$1.17^{+0.15}_{-0.16}$	$0.33^{+0.07}_{-0.07}$	14.37	0.82	$0.19^{+0.02}_{-0.02}$	0.33	$0.24^{+0.09}_{-0.04}$	$0.10^{+0.02}_{-0.02}$	$0.26^{+0.09}_{-0.05}$
Dynamical ($r_{\mathrm{rms}}\sigma^2$)	$0.94^{+0.14}_{-0.15}$	$0.39^{+0.09}_{-0.08}$	8.80	0.86	$0.08^{+0.01}_{-0.01}$	0.30	$0.28^{+0.08}_{-0.05}$	$0.10^{+0.03}_{-0.02}$	$0.30^{+0.09}_{-0.05}$
Dynamical (σ^3)	$1.05^{+0.15}_{-0.17}$	$0.35^{+0.07}_{-0.07}$	8.48	0.84	$0.12^{+0.01}_{-0.01}$	0.31	$0.28^{+0.08}_{-0.05}$	$0.10^{+0.03}_{-0.02}$	$0.29^{+0.09}_{-0.05}$
Radius	$2.33^{+0.41}_{-0.47}$	$0.21^{+0.09}_{-0.09}$	3.04	0.84	$0.04^{+0.004}_{-0.004}$	0.35	$0.33^{+0.10}_{-0.05}$	$0.12^{+0.03}_{-0.02}$	$0.35^{+0.10}_{-0.05}$

Notes. [a]Spearman rank correlation. All results have >95 per cent significance.
[b]Estimated fractional scatters in proxy and mass in dex.
[c]Intrinsic scatter in mass accounting for statistical scatter in the proxy (see equation 17).
[d]The systematic scatter in mass, $\sigma^2_{\log M,\mathrm{sys}} = \sigma^2_{\log M,\mathrm{int}} + \sigma^2_{\log M,\mathrm{cal}}$.

selection methods these mass proxies are highly correlated with M_{500} with a low degree of scatter about the regression.

Comparing the volumetric $\tilde{N}_{1\,\mathrm{Mpc}}$ against the literature, we note that Budzynski et al. (2012) compute a similar relation for high mass clusters ($M_{500} > 10^{13.7}\,\mathrm{M}_\odot$) and find that $M_{500} \propto N_{1\mathrm{Mpc}}^{1.4\pm0.1}$. This relation is shallower than ours, though it is (just) consistent within our 1σ error. Conversely, our FoF result is consistent with a 1:1 relation between mass and richness as one would expect if SFE and galaxy LF were independent of halo mass.

Likewise, we compare the $L_{1\,\mathrm{Mpc}}$ result to the result of Popesso et al. (2007) who find an L_{200}–M_{200} relation with a slope of 0.92 ± 0.03. Converting this to an M_{500}–$L_{1\mathrm{Mpc}}$ relation using the NFW model discussed below gives a slope of $\sim$1.5, which is well within our estimated error. As with FoF richness, the FoF luminosity is also consistent with a slope of unity. A noticeable kink in the volumetric mass–luminosity data is apparent at $\sim 10^{14}\,\mathrm{M}_\odot$. There is some indication of a similar feature in a number of the other mass–proxy relations, though as it is not seen in any proxy–proxy relation and given the limited sample size it is hard to be confident of its reality.

We note from Table 3 that there is a significant difference between the slopes of the volumetric and FoF samples, indicating that, as one would expect, selection plays a significant role. This difference can be plausibly attributed to the use of a 1 Mpc aperture for the volumetric sample. For example, at low mass, where $r_{500} \ll 1$ Mpc this would result in recovering $N_{1\,\mathrm{Mpc}} > N_{500}$, and similarly recovering $N_{1\,\mathrm{Mpc}} < N_{500}$ at high mass. Conversely, a Friends-of-Friends analysis should link galaxies together in a way that scales with the size of the group. Hence, we expect the volumetric scaling relations to be steeper.

To investigate further, we construct a simple analytic model. We assume that galaxies are distributed as an NFW with a concentration half that of the dark matter (Budzynski et al. 2012), and take the dark matter concentration from the mass–concentration relation derived from simulations of relaxed haloes by Duffy et al. (2008):

$$c_{200} = 6.71 \pm 0.12 \left(\frac{M_{200}}{2\times10^{14}\,h^{-1}\,\mathrm{M}_\odot}\right)^{-0.091} (1+z)^{-0.44}. \quad (19)$$

We assume $z = 0.05$, close to the mean of our sample, and find that including scatter in the mass–concentration relation has a negligible impact on our results, so we do not include it. We then model the aperture richness N_{ap} as

$$N_{\mathrm{ap}} = \int_0^{1\,\mathrm{Mpc}} 2\pi r\,\Sigma(r, r_{\mathrm{s}}(M_{200}))\mathrm{d}r, \quad (20)$$

where Σ is the projected NFW from equation (7) and $r_{\mathrm{s}} = 2r_{200}/c_{200}$. We then calculate N_{ap} for a set of equally spaced $\log M_{200}$ values in the range $12.5 \lesssim \log_{10}(M_{200}/\mathrm{M}_\odot) \lesssim 15.5$. We also convert each M_{200} to M_{500} using the 3D NFW, equation (6), and concentration c_{200}.

Applying a BCES (X|Y) regression to the resulting points in the M_{500}–N_{ap} plane gives a slope of 1.42 ± 0.02. This is shallower than our observational result, though marginally within the errors. Whilst the effect of an aperture does indeed appear to steepen the relation, this suggests that a second effect may be at work.

A possible second source of discrepancy could be variation of SFE as a function of halo mass (see e.g. Behroozi, Conroy & Wechsler 2010; Moster et al. 2010; Leauthaud et al. 2012; Budzynski et al. 2014). This variation is both inferred through abundance matching techniques (e.g. Behroozi et al. 2010) and measured directly using, for example, cluster mass-to-light ratios (e.g. Budzynski et al. 2014). This would boost the stellar mass fraction in lower mass haloes relative to those of higher mass. Using the Budzynski et al. (2014) stellar mass fraction–M_{500} relation, and comparing it to a global stellar mass fraction from the SDSS, $\Omega_* = 1.9 \times 10^{-3}$ (Panter et al. 2007), we can determine a mass-dependent bias factor to scale the modelled M_{500}–N_{ap} relation discussed previously. As our groups are an order of magnitude more massive than the halo

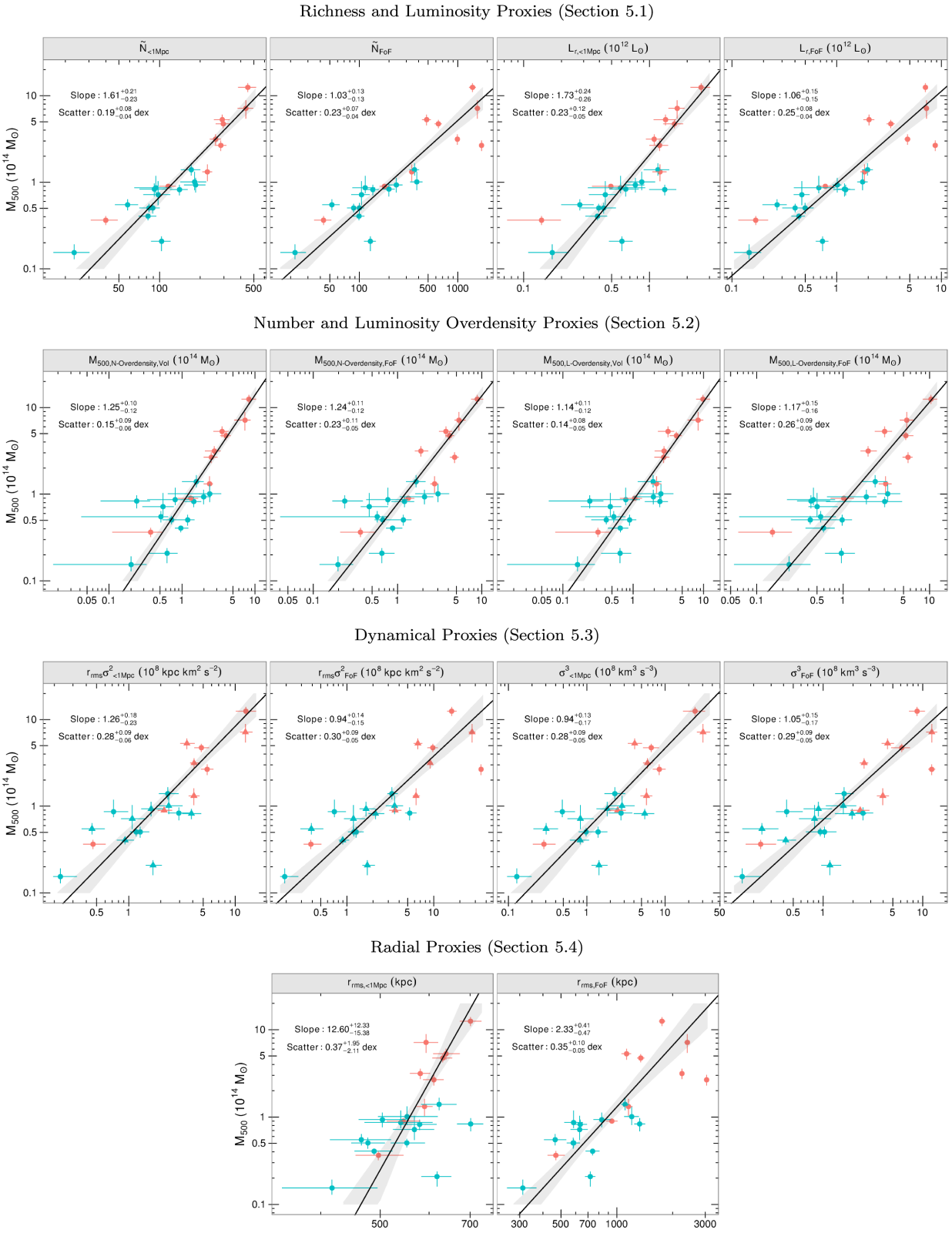

Figure 2. X-ray M_{500} plotted against each mass proxy with, from left to right, top to bottom, richness, luminosity, number overdensity, luminosity overdensity, $r_{\rm rms}\sigma^2$ dynamical, σ^3 dynamical and $r_{\rm rms}$ radius. Alternating plots show results for the volumetric and FoF sample. Blue points are from the S09 sample, red points from the SP10 group and cluster samples, respectively (see Table 1). The solid lines and shaded regions show the BCES(X|Y) regressions and a 68 per cent interval drawn from the distribution of bootstrapped fit results. For the dynamical estimators (third row), groups with a large fraction of central galaxies with 'cloned' redshifts (see Section 2.2.3) are shown by triangular points.

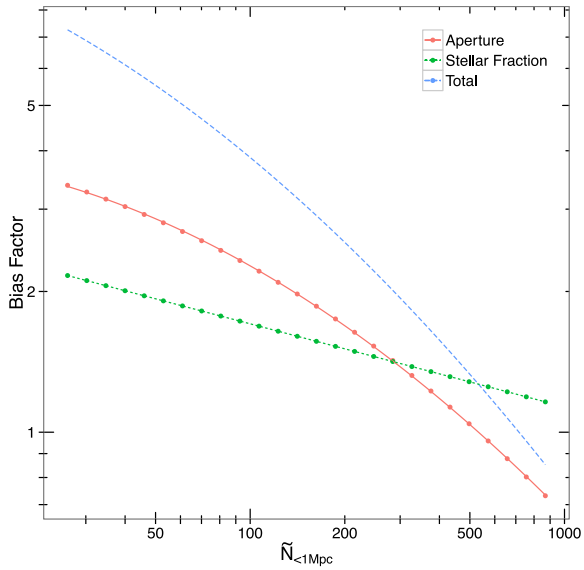

Figure 3. Modelled bias factors and best-fitting curves. These are required to rescale $\tilde{N}_{1\,\mathrm{Mpc}}$ to N_{500} (C_N, red, solid line) and to account for variation in SFE (C_*, green, dotted line). The combined scale factors, ($C_N C_*$) are shown as the blue, dashed line.

mass where the stellar mass fraction peaks ($M_{500} \sim 10^{12}\,\mathrm{M}_\odot$), in the mass regime where the change in stellar mass fraction is relatively shallow, we would expect to see limited variation in SFE. However, this small variation appears sufficient to modify the slope of the modelled M_{500}–N_{ap} relation to 1.67 ± 0.03 – close to our observed relations for aperture richness and luminosity.

We can cast these two effects into two bias factors, $C_N = N_{\mathrm{ap}}/N_{500}$ and $C_* = f_*/\bar{f}_*$, where $\bar{f}_* = \Omega_*/\Omega_{\mathrm{m}}$. In Fig. 3, we show the magnitude of these bias factors as a function of $N_{1\,\mathrm{Mpc}}$.

Unlike the volumetric sample, the richness and luminosity proxies for the FoF sample require no aperture correction. We attempt instead to correct N_{FoF} and L_{FoF} for SFE only. Scaling for C_* brings the observed slopes down to 0.92 ± 0.10 and 0.96 ± 0.11, respectively, both consistent with, though rather lower than, the expected slope of unity. We also note that even without this factor, both FoF-selected proxies are already consistent with unity. This may relate to the nature of the FoF algorithm used. As we will discuss in Section 6, we believe that this is rather overgenerous in terms of linking galaxies together. If this effect is stronger for higher mass systems it would lead to some flattening of the observed mass–proxy relation.

We perform one final check on our data by examining the luminosity–richness relation for our sample. Popesso et al. (2007) found luminosity and richness to be well correlated, with a slope of 1.00 ± 0.03. Whilst we use a slightly different definition of both group richness and luminosity, we also find (Fig. 4) that they are strongly correlated (Spearman rank correlations of 0.95 and 0.99 for volumetric and FoF samples, respectively). BCES orthogonal regression slopes are $0.96^{+0.07}_{-0.08}$ and 0.98 ± 0.03 for our two selection methods, both in excellent agreement with each other and consistent with the Popesso et al. (2007) result. This suggests that there is no systematic trend in mean galaxy luminosity with system mass within our sample.

5.1.1 *Stellar mass*

As group stellar mass is a more fundamental property than optical luminosity, one might expect stellar mass to also be a useful mass

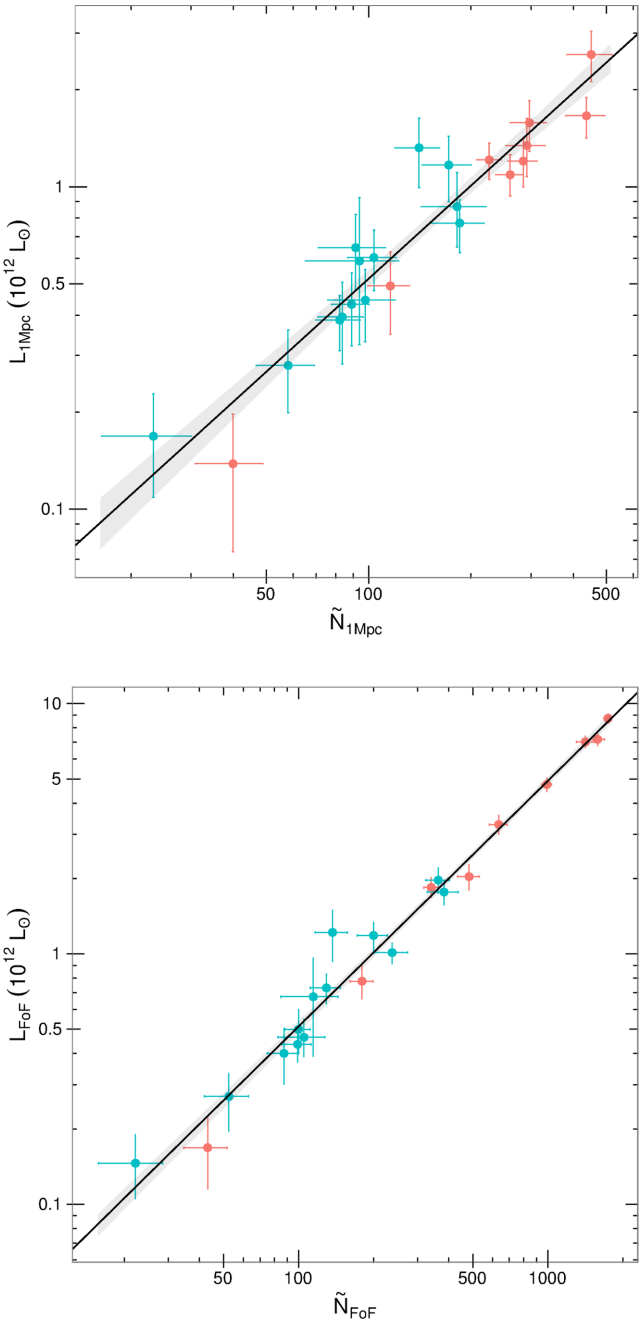

Figure 4. Top: background-subtracted optical luminosity against richness within 1 Mpc. Bottom: total FoF luminosity against FoF richness. Colour coding is as for Fig. 2.

proxy. In practice, however, with photometric data typical of optical surveys, the stellar mass has to be estimated from the luminosity and colour of member galaxies. By using such a colour-dependent mass-to-light ratio (e.g. Bell et al. 2003), estimates of the group stellar mass are a more indirect halo mass proxy than luminosity or richness. We nevertheless briefly investigate the effectiveness of such an approach.

To estimate group stellar mass we use the mass-to-light ratio of Bell et al. (2003), $\log_{10}(M/L_r) = -0.306 + 1.097(g - r) - 0.1$ (see also Budzynski et al. 2014). We cut our galaxy sample further to include groups that are complete to an absolute magnitude $M_r < -20$ (equivalent to groups extending to $z = 0.07$). Using this

truncated sample of 18 groups, we calculate a luminosity-weighted mean $\log_{10}(M/L_r)$ for each group. We estimate the stellar mass of the halo as the product the total luminosity of each group for galaxies brighter than $M_r < -20$ and the characteristic mass-to-light of that group. As with other estimators, the volumetric sample repeats this using the 3–5 Mpc aperture to estimate a background contribution which is then subtracted from the stellar mass measured within 1 Mpc. No background correction is applied to the FoF sample.

This brief examination shows halo mass–stellar mass relations that are comparable to the luminosity estimates, with slopes of $1.83^{+0.29}_{-0.37}$ and $1.10^{+0.17}_{-0.16}$ for the volumetric and FoF samples, respectively. Interestingly, we observe a marginally larger degree of intrinsic scatter in the volumetric sample ($0.24^{+0.15}_{-0.06}$ dex) than that observed for the luminosity-based proxy. Including calibration uncertainties similarly produces a greater degree of systematic uncertainty of $0.27^{+0.14}_{-0.06}$ dex.

Unfortunately, the bootstrap analysis of the FoF sample results in large statistical uncertainties on stellar mass. These large uncertainties, possibly due to large interloper populations degrading the mass-to-light estimation, are too large to allow the systematic uncertainty on the FoF stellar mass based proxy to be constrained. We note however that the total scatter observed is again larger than for the FoF group luminosity at 0.28 dex.

The implication is that whilst stellar mass can be used as a halo mass proxy, the assumptions made in converting luminosity to stellar mass can introduce substantial uncertainty. For this reason, this study favours the use of luminosity over stellar mass as a halo mass proxy. We also note that whilst we can empirically estimate the variation in SFE, the large uncertainty we find again causes us to prefer the Budzynski et al. (2014) stellar mass fraction formalism.

5.2 Overdensity

The second row of Fig. 2 shows the observational results of the number and luminosity-based overdensity analyses, with fit parameters given in Table 3. The number overdensity mass proxy performs well, exhibiting a strong correlation with mass and the lowest levels of intrinsic and calibration scatter for both volumetric- and FoF-selected groups. However, it also has the largest statistical errors on the measured proxy value.

As this proxy provides a direct measure of group mass, we would expect the mass–proxy relation to have a slope of unity. We observe however that whilst our richness overdensity masses are broadly consistent with X-ray masses, the slope of the relation is significantly steeper than unity, implying that there is a mass-dependent bias in the measured overdensities. A similar, though slightly less significant, bias is observed for the luminosity overdensity. The reduced bias for luminosity may be due to its reduced sensitivity to the faint-end slope of the LF.

A fundamental assumption in the use of the galaxy (and galaxy luminosity) overdensity is that galaxies are a perfect tracer of the underlying dark matter. As noted previously however, it is well known that galaxy density profiles have concentrations different to that of their dark matter counterparts (see e.g. Hansen et al. 2005; Budzynski et al. 2012). However, this would be expected to make only a minor difference to the integrated galaxy density within r_{500}. More significant is likely to be the bias discussed in Section 5.1, which would arise from halo mass dependence in the efficiency of star formation. Some bias may also be introduced by the comparison between our observed mean density and the global mean density, which incorporates an assumption about the stellar mass-to-light ratio.

It is well known that clusters and groups have different stellar mass-to-light ratios compared to the Universe as a whole, a property that varies with the scale of the system (see e.g. Ostriker, Peebles & Yahil 1974; Davis et al. 1980). We investigate this briefly here, again using the colour-dependent mass-to-light ratio of Bell et al. (2003). Using the g and r absolute magnitudes from the NYU-VAGC catalogues from which our sample was drawn in Section 2, we find that the mean log mass-to-light ratio of the global sample in the range $M_r < -19$ and $0.01 \leq z \leq 0.1$ gives a mass-to-light ratio of 2.8. Repeating this for our clusters, using galaxies projected within 1 Mpc, we find a mean mass-to-light ratio of 3.6. This implies that our overdensity analysis would systematically underestimate the overdensity of haloes by a factor of 1.3. In practice, the mass-to-light ratio of clusters varies with halo mass (e.g. Budzynski et al. 2014) so we also estimate the mass-to-light ratio of our groups in mass bins split at group masses of $M = 7.5 \times 10^{13}$ and $2 \times 10^{14}\,\mathrm{M}_\odot$. We do not see any substantial change with group mass (mean mass-to-light values of 3.5, 3.6 and 3.8, respectively) so we adopt the mean factor above. This translates to underestimating masses by $\sim$1.5, bringing our relation into agreement with expectations at high mass ($\sim 10^{15}\,\mathrm{M}_\odot$). However, as this correction is mass independent it will have no effect on the measured slope.

As with Section 5.1, we again look to variation in SFE and the consequent stellar mass fraction of haloes as an effect that will modify the slope of our observed relation. This is also established to be different in groups and clusters compared to global values (see e.g. Behroozi et al. 2010; Moster et al. 2010; Leauthaud et al. 2012; Budzynski et al. 2014). It is also known to be mass dependent (e.g. Leauthaud et al. 2012; Budzynski et al. 2014), such that low-mass haloes have higher stellar mass fractions compared to high-mass systems, which are considerably closer to the global value. A halo with higher SFE will have a higher density of galaxies and light than might be expected for its mass overdensity (ignoring galaxy mergers). A correction for this would involve scaling the modelled halo densities by a factor that would allow a correct comparison to the global density. We define such a factor as $C_O = f_*(M_{500})/\bar{f}_*$ where $\log_{10}(f_*(M_{500})) = -0.11 \log_{10}(M_{500}/3 \times 10^{14}\,\mathrm{M}_\odot) - 2.04$ (Budzynski et al. 2014) and $\bar{f}_* = \Omega_*/\Omega_\mathrm{m}$. Again, we use $\Omega_* = 1.9 \times 10^{-3}$ derived from the SDSS (Panter et al. 2007).

Applying a correction for both a variable SFE and the difference between global and cluster mass-to-light ratios, we find that the overdensity mass–X-ray mass relations are consistent with our expectations. A richness-based overdensity recovers slopes of $1.03^{+0.08}_{-0.09}$ and 1.00 ± 0.10 for the volumetric and FoF samples, respectively, with no significant bias in normalization. Luminosity similarly is corrected to a slope of $0.94^{+0.08}_{-0.10}$ (volumetric) and $0.96^{+0.12}_{-0.14}$ (FoF). Given that we now understand the discrepancy between observations and expectations, we believe that it is safe to simply calibrate the uncorrected overdensity mass–X-ray mass relation.

5.3 Dynamical mass

Results for the dynamical mass estimators are shown in the third row of Fig. 2 and in Table 3. The 12 groups for which there is a central population of galaxies with cloned redshifts due to fibre collisions, as discussed in Section 2.2.3, are flagged in these figures, but do not appear to be systematically offset. Excluding these systems from the fit results in only modest changes to the fitted slope (to $\alpha = 1.12$ and 0.85 for the volumetric $r_\mathrm{rms}\sigma^2$ and σ^3 relations, respectively).

These changes lie within the statistical errors, and we therefore retain the affected systems in our analysis.

Whilst the two proxies for both selection methods appear well correlated with the X-ray masses, they also show a large amount of scatter compared to some of our other mass estimators. From the Virial Theorem, we expect these estimators to scale linearly with mass. We find that all but the volumetric $r_{\rm rms}\sigma^2$ estimator are consistent with this expectation at the 1σ level. Given the highly discrepant behaviour of the $r_{\rm rms}$ proxy for the volumetric sample (discussed in Section 5.4), the behaviour of the $r_{\rm rms}\sigma^2$ mass proxy is not surprising.

We note that our three good dynamical estimators – two FoF and one volumetric – have mass–proxy slopes which are consistent with each other, indicating that the extracted velocity dispersion is quite robust against the treatment of interlopers and the use of apertures. We also see no strong indication of the biased velocity dispersions at low mass suggested by previous work (e.g. Osmond & Ponman 2004). However, we note that our sample has been selected to be X-ray bright and morphologically relaxed, whilst the Osmond & Ponman (2004) study involved a more diverse set of galaxy groups.

5.4 Radii

The observed rms radius–mass relation is shown in the final row of Fig. 2, with fit parameters in Table 3. Based on $M \propto r^3$ at constant density (in this case, overdensity), we would expect the mass–radius relation to have a slope of 3. We instead find that the relation for the volumetric sample is much steeper, whilst the slope for the FoF sample is somewhat shallower than expected by almost 2σ. This mass proxy is the worst performing of those discussed in this paper with the largest final systematic uncertainty of both samples.

The difference in the fitted relations between the volumetric and FoF selection methods is likely to be due to the use of a metric aperture for the volumetric sample, coupled with the impact of interlopers. Interlopers, due to their uniform distribution, would be expected to increase the measured rms radii, especially for poorer groups, steepening the observed slope. At the same time, the imposition of an aperture sets an upper limit to the radius. The result is a very small dynamic range in $r_{\rm rms}$, as can be seen in the bottom-left panel of Fig. 2.

To investigate these effects, we construct a model to probabilistically sample projected NFW models, constructing haloes of a given richness, and estimating their $r_{\rm rms}$ radii. We base our approach on the method Budzynski et al. (2012) used to estimate mass completeness. We first estimate the number of galaxies contributed by a halo of a given mass using our observed mass–aperture richness relation. This relation is calibrated to recover galaxies brighter than a fixed absolute magnitude, $\tilde{N}_{1\,\rm Mpc}$, whilst we require the number of galaxies that would actually be observed at some specific redshift, $N_{1\,\rm Mpc}$. We account for this by removing the correction given by equation (4), adopting a redshift of $z = 0.05$ for our simulated group. A projected NFW with a galaxy concentration as described in Section 5.1 is then normalized to recover this halo richness. We add a background term to introduce interloper contamination and integrate the combined normalized NFW and interloper density profile within our aperture to estimate the 'observed' aperture richness.

To probabilistically construct a group with a halo of given mass, we draw the desired number of galaxies (N) from a Poisson distribution using the estimated richness as the mean. We then draw $N - 1$ radii from the projected NFW likelihood function, equation (10), and add a single galaxy to the centre for the BGG. We generate 50 000 such simulated groups uniformly spanning the mass range

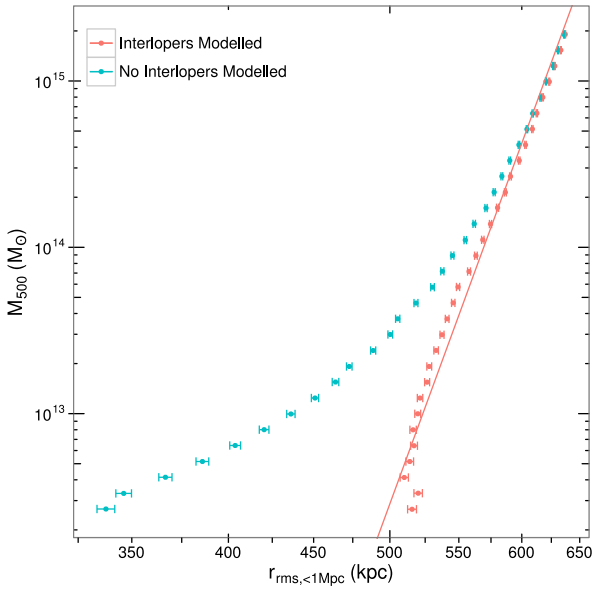

Figure 5. Red: modelled $r_{\rm rms}$ radius including interloper contamination with BCES(X|Y) best fit line shown. Blue: modelled $r_{\rm rms}$ radius not including interloper contamination.

$12.5 \leq \log_{10}(M_{200}/{\rm M}_\odot) \leq 15.5$. We generate our simulations for two scenarios: one where we include no interloper contamination, and one where we assume an interloper population with galaxy number density $(1.0 \pm 0.3) \times 10^{-6}$ kpc^{-2}, based on the mean background measured by our overdensity method (Section 5.2).

Finally, we bin these two sets of 50 000 randomizations into bins of 0.1 dex in mass. The resulting mass–radius relations are shown in Fig. 5. The small dynamic range in $r_{\rm rms}$ is similar to that seen in Fig. 2, and power-law fits to the sample including interlopers gives a slope of 27.4 ± 1.1 – far steeper than the naive expected value of 3. It can be seen that the inclusion of an interloper population does affect the recovered slope, especially in the lower mass regime where the 'clean' sample diverges significantly from the interloper contaminated model. However, the clean sample also appears to approach a slope of 3 at low $r_{\rm rms}$, equivalent to groups with $r_{500} \ll 1$ Mpc. Our conclusion is that both interlopers and aperture effects play a significant role in biasing this relation, with interlopers dominating at low masses, increasing the rms radius of a group, whilst the aperture dominates at high masses where it truncates the galaxy distribution, reducing the rms radius. Ultimately, the derived radius for the volumetric selection is biased to the point where we observe a near constant $r_{\rm rms}$, making it unsuitable for use as a mass proxy. As an aperture is not used for the FoF sample, interloper contamination will be the main source of bias at all masses.

5.5 Uncorrected effects

Throughout this section we have discussed mass-dependent biases that can be introduced by failing to take into account properties of the sample selection (such as aperture effects) or of the groups themselves (such as variation in SFE). We have not, however, discussed the possible biases introduced by assumptions we have made. The most important is the implicit assumption, at various points in our analysis, of a universal galaxy LF within our systems. In practice, it has been observed (e.g. Hansen et al. 2005; Robotham et al. 2006, 2010; Zandivarez & Martínez 2011) that the group LF varies as a function of halo mass, becoming similar to the global galaxy LF

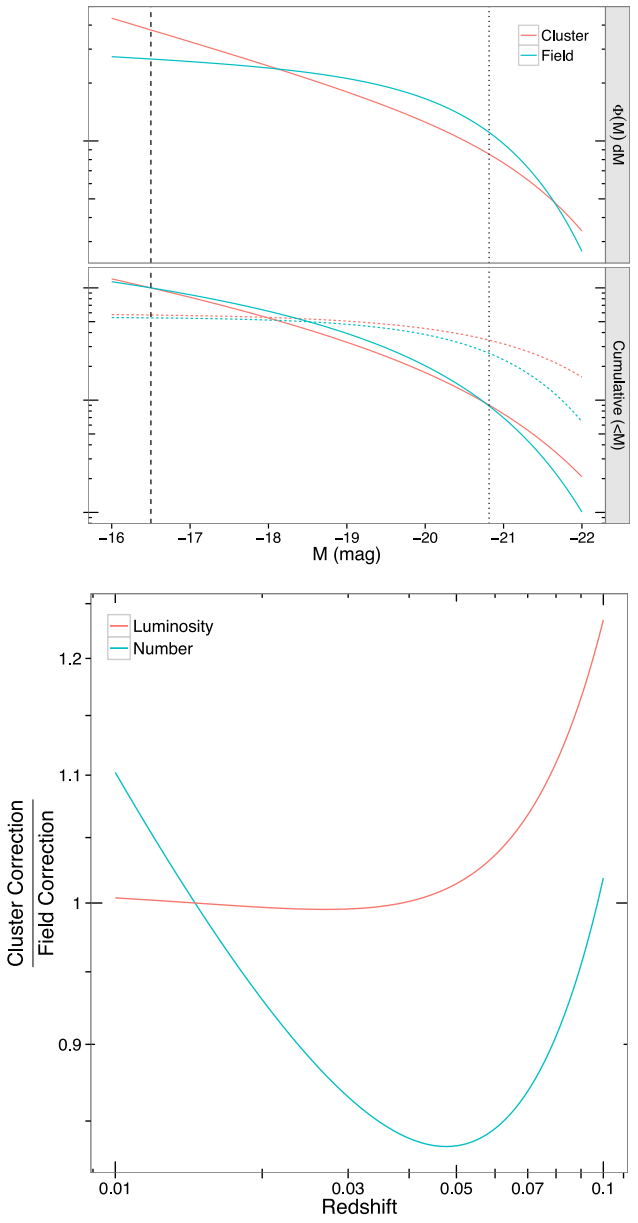

Figure 6. Top, upper panel: LFs used for the cluster correction (red curve) and the global LF (blue curve) with arbitrary normalization. Top, lower panel: the cumulative number (solid coloured line) and luminosity (dashed coloured lines) density of the LFs above. The vertical lines illustrate the magnitude we extrapolate to ($M_r = -16.5$) and the absolute magnitude corresponding to the flux limit at $z = 0.1$, $M_r = -20.8$. Bottom: the degree to which a high-mass cluster LF would overcorrect the richness and luminosity of a low-mass group as a function of redshift. This assumes that low-mass groups have a LF comparable to a global LF.

at low mass. As we use a single LF for completeness correction, this has the potential to bias low-mass systems whose LF differs significantly from the Popesso et al. cluster LF used. This will affect both luminosity and richness estimators and will be a larger effect for cases where a significant degree of extrapolation is required, i.e. at higher redshift. Whilst we do not attempt to correct for this, under the assumption that low-mass groups more closely resemble the field (i.e. global) LF of Blanton et al. (2003), we can examine the degree of overcorrection that extrapolating with an inappropriate LF can cause as a function of redshift. This is shown in Fig. 6.

These differences are caused by the cluster LF featuring a brighter turnover magnitude than the field [$M_* - 5\log_{10}h = -21.35$ (Popesso et al. 2005) rather than -20.44 (Blanton et al. 2003)] and a steeper faint-end slope ($\alpha = -1.30$ compared to -1.05). When little extrapolation is required (at low redshift where only a small portion of the faint end is lost due to the survey limit), the low luminosities of galaxies below the survey flux limit result in little discrepancy between a cluster- and field-based luminosity correction, whilst their greater abundance in the field results in a cluster based richness correction underestimating rescale factors by $\sim$10 per cent. Both factors converge to no discrepancy at $z \approx 0.015$ where the limiting magnitude is comparable to our adopted magnitude limit of $M_r = -16.5$. Conversely, at higher redshifts where it is necessary to extrapolate from near M_* ($M_{\rm cut} \approx -20.8$ at $z = 0.1$), the excess of bright galaxies from the cluster LF cause both the number and luminosity corrections to rise.

Hence, if our low-mass groups have LFs similar to the field, then attempting to correct them for incompleteness in faint galaxies using a cluster LF will have two main effects as a function of redshift. At low redshift, the degree of overcorrection will be negligible for a luminosity correction, rising to >20 per cent at $z = 0.1$. Conversely, richness corrections will be underestimated by 10–20 per cent until $z \sim 0.1$ at which point they become overestimated and continue to rise with increasing redshift. Correcting for these effects would require knowledge of the way in which LFs vary as a function of system mass. In practice, our fitted mass–proxy relations should calibrate out this effect, at least for the relaxed X-ray bright groups which constitute our sample.

Similarly, we have also not discussed the impact of galaxy mergers, which will affect the richness and luminosity estimators, as richness is reduced and stars are stripped to form the ICL. If these effects are mass dependent, we assume their average effects can be calibrated out in our analysis. However, the varying merger histories of groups will also introduce scatter into the relevant mass–proxy relations.

It is important to emphasize that this study is based on a low-redshift sample, and apart from inclusion of the redshift dependence of $\rho_{\rm crit}$ (which is small over the redshift range of our sample) when calculating r_{500} for our systems, we do not take into account the evolution of mass proxies with redshift. Redshift correction of the dynamical and radius estimators is straightforward, and follows directly from the evolution in $\rho_{\rm crit}$. The same is true of the overdensity estimators, provided that evolution in galaxy luminosities is the same in clusters as in the field. However, we would expect the richness and luminosity mass proxies to be strongly dependent on the evolution of the galaxy LF within galaxy groups and clusters. Recent work implies that there may be little evolution in the mass–richness proxy (Andreon & Congdon 2014), though as this work uses a specific subset of galaxies to define richness, the result cannot necessarily be extrapolated to our richness proxy.

Finally, it should also be noted that our mass–proxy relations have been derived from an X-ray-selected sample. Since the hot intergalactic gas responsible for the bulk of the X-ray emission is heated during the collapse of the group halo, X-ray-bright groups represent a subset of the broader group population. Groups selected in this way have the advantage of avoiding the inclusion of spurious groups, arising from line-of-sight superposition of unrelated galaxies; however, the groups in our sample will be more evolved and relaxed than many groups in optically selected samples. This is seen clearly in the studies of Rasmussen et al. (2006) and Balogh et al. (2011) which found optically selected samples to contain many groups with little or no X-ray emission. This could result

Table 4. Spearman rank correlations, with bootstrap errors, for the fractional offsets of masses predicted for (upper triangle) each volumetric-selected proxy, and (lower triangle) FoF-selected proxy from the X-ray mass.

	Richness	Luminosity	Overdensity (N)	Overdensity (L)	Dynamical ($r_{\rm rms}\sigma^2$)	Dynamical (σ^3)	Radius
Richness	–	$0.78^{+0.14}_{-0.09}$	$0.86^{+0.09}_{-0.04}$	$0.81^{+0.11}_{-0.06}$	$0.69^{+0.17}_{-0.13}$	$0.71^{+0.16}_{-0.11}$	$0.24^{+0.25}_{-0.22}$
Luminosity	$0.91^{+0.09}_{-0.05}$	–	$0.60^{+0.18}_{-0.13}$	$0.65^{+0.18}_{-0.13}$	$0.68^{+0.19}_{-0.12}$	$0.64^{+0.20}_{-0.14}$	$0.56^{+0.20}_{-0.15}$
Overdensity (N)	$0.55^{+0.21}_{-0.17}$	$0.53^{+0.21}_{-0.16}$	–	$0.89^{+0.10}_{-0.05}$	$0.59^{+0.21}_{-0.19}$	$0.64^{+0.19}_{-0.16}$	$-0.03^{+0.25}_{-0.26}$
Overdensity (L)	$0.56^{+0.20}_{-0.15}$	$0.66^{+0.19}_{-0.13}$	$0.89^{+0.08}_{-0.04}$	–	$0.60^{+0.22}_{-0.18}$	$0.65^{+0.20}_{-0.15}$	$-0.03^{+0.26}_{-0.25}$
Dynamical ($r_{\rm rms}\sigma^2$)	$0.84^{+0.12}_{-0.08}$	$0.80^{+0.11}_{-0.06}$	$0.46^{+0.23}_{-0.20}$	$0.46^{+0.22}_{-0.18}$	–	$0.98^{+0.03}_{-0.01}$	$0.49^{+0.22}_{-0.17}$
Dynamical (σ^3)	$0.66^{+0.19}_{-0.15}$	$0.64^{+0.18}_{-0.13}$	$0.52^{+0.20}_{-0.18}$	$0.54^{+0.19}_{-0.18}$	$0.90^{+0.10}_{-0.07}$	–	$0.37^{+0.24}_{-0.19}$
Radius	$0.91^{+0.08}_{-0.03}$	$0.78^{+0.15}_{-0.10}$	$0.32^{+0.26}_{-0.22}$	$0.28^{+0.24}_{-0.21}$	$0.42^{+0.22}_{-0.18}$	$0.31^{+0.24}_{-0.20}$	–

from lack of virialization, or from the effects of strong feedback raising the gas entropy (e.g. Voit & Bryan 2001). Such effects could potentially impact on some of our mass proxies. For example, unvirialized systems cannot be expected to provide reliable dynamical mass estimates. This emphasises the value of employing a number of mass estimators, rather than relying on a single method.

5.6 Discussion

In general, the majority of our proxies exhibit mass–proxy relations that are either consistent with theoretical expectations, or for which any discrepancy can be plausibly explained. The steepened slopes of volumetric aperture richness, luminosity and rms radius proxies are consistent with the expected effects of imposing an aperture, and the behaviour of richness, luminosity and overdensity proxies are also influenced by known variations in SFE. Our approach is to calibrate out these effects in deriving our proxy–mass scaling relations.

What do our results suggest is the *best* mass proxy from the set we have studied? One reasonable definition of 'best' is the proxy which has the smallest systematic error (arising from intrinsic population scatter and calibration error). Reference to the values of $\sigma_{\log M,\,\rm sys}$ listed in Table 3 indicates that the volumetric overdensity mass proxies constitute our best mass proxies. These proxies may be the most reliable estimators when they can be measured to high precision, but it can be seen from the same table that they also have the largest *statistical* errors ($\sigma_{\log x,\,\rm stat}$) of any of our proxies.

In the absence of a high-quality galaxy surface density profile, we find that a simple richness estimator provides the most reliable result, with ∼0.2 dex (i.e. ∼50 per cent) systematic scatter and relatively small statistical errors for galaxy samples of the size ($\gtrsim 15$) used in this study. This is true for both volumetric and FoF selection methods.

However, we believe that there is substantial benefit in employing not just a single mass estimator, but a basket of them. We will return to this point in Section 8, but an indication of the extra information provided by using different mass proxies in conjunction can be gleaned by examining the *residuals* about our regression fits for the different proxies. Specifically, we look for evidence of correlations between residuals of different proxies, which might, for example, imply that if a group's mass is overpredicted by one proxy it will also be overpredicted by another.

Of course, it is self-evident that some proxies will be intrinsically correlated. For example, richness and luminosity, or the two dynamical estimators. However, others – such as luminosity and σ^3 – would appear to be quite independent of one another, apart from the relationship of both to system mass. We define residuals

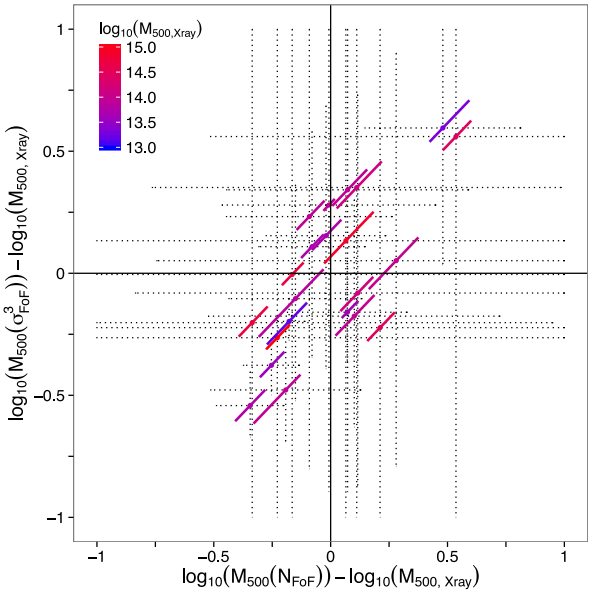

Figure 7. Fractional residuals about the regression lines for the FoF richness and σ^3 dynamical estimators defined as $R = \log_{10}(M_{500}(x)) - \log_{10}(M_{500,\,\rm X})$. Bold, solid lines: the 1σ range due to statistical errors on the measured X-ray masses. Dotted lines: 1σ range due to the statistical errors on the mass proxies, intrinsic scatter in the relation and calibration errors added in quadrature. Each data point is coloured according to its X-ray mass, with blue $\sim 10^{13}\,{\rm M}_\odot$ and red $\sim 10^{15}\,{\rm M}_\odot$.

as the log-space difference between a mass predicted by the proxy (x) and the 'true' X-ray mass, $R_x = \log_{10}(M_{500}(x)) - \log_{10}(M_{500,\,\rm X})$, and test for correlation between residuals from the different mass proxies using the Spearman rank statistic. The results are shown in Table 4 for both selection methods with errors derived from bootstrap resampling our residual estimates. Note however that the rms radius proxy is intrinsically flawed for volumetric selection.

Surprisingly, the residuals are positively correlated in most cases. To consider what this means, we focus on the correlation between the richness and σ^3 mass estimators. The residuals for these two are correlated with $\rho = 0.66^{+0.19}_{-0.15}$ and $0.71^{+0.16}_{-0.11}$ for the FoF and volumetric samples, both of which are significant at >95 per cent with errors, again, from a bootstrap analysis. We plot the residuals from the fitted, FoF mass–proxy relations for these two mass estimators in Fig. 7. Here a group which lies on both calibration relations would lie at the origin.

The clear correlation between the residuals indicates that for a group in which the richness proxy produces a mass overestimate (relative to its measured X-ray mass), the same tends to be true

for the dynamical mass estimator, and similarly for underestimates. The colour coding in the figure indicates the temperature of each system, and no correlation between this and location on the plot is apparent. There are two possibilities: either perturbations from the mean mass relations in richness and dynamics are related in some way, or there are errors in the X-ray masses. The latter would induce correlated offsets of the type we observe – a group with an overestimated X-ray mass would fall in the lower-left quadrant, whilst an X-ray underestimate would move it towards the top right.

The fact that the *vast majority* of our varied mass estimators show positively correlated residuals (see Table 4) suggests strongly that the X-ray masses are at fault. The alternative would be some sort of conspiracy between the whole set of mass proxies to perturb together. Of course there are *statistical* errors on the X-ray mass measurements, and the effect of these is shown (diagonal bars) in the figure. These would introduce some elongation of the distribution in the direction observed, but it is clear that in about a third of the systems these statistical errors cannot explain the magnitude of their offset from the centre. So, there is evidence here for errors in the X-ray-derived masses, over and above the estimated statistical errors, in a significant subset of our group sample. For the worst cases, errors in the X-ray mass of a factor of 2 or more (in both directions) are implied.

The dotted error bars on the plot, in the x and y directions represent estimates of the error on the proxy mass arising from a combination of the measurement error in the proxy quantity, the calibration error in converting it to a mass and our estimate of the intrinsic scatter in the proxy value across the population (see Section 4.2). It is immediately apparent that these error bars are far too large – i.e. they are inconsistent with the observed scatter in the data. Table 3 shows that this error bar is dominated by the estimated contribution ($\sigma_{\log M,\,\mathrm{int}}$) from scatter in the population. This in turn has been estimated (equation 17) by subtracting the expected X-ray mass error and the propagated error in the proxy measurement from the observed scatter in mass about the regression relation. If there are offsets in the X-ray masses over and above their statistical errors, which it appears that there are, then $\sigma_{\log M,\,\mathrm{int}}$ will be overestimated. It seems clear that this is the case.

Our analysis thus far has made use of X-ray masses derived from a hydrostatic mass analysis. However, a simpler X-ray mass estimate can be derived from the mean X-ray temperature, using a mass–temperature relation. We have checked that the overall picture seen in Fig. 7 remains unchanged if such temperature-based masses are used. The most discrepant groups in the figure, remain similarly offset with the alternative X-ray masses.

6 MOCK GROUPS

The results presented above are derived from a sample of relaxed X-ray bright systems. In practice, it would be very useful to have mass estimates for a wider range of groups, such as that generated by optical selection from a galaxy redshift survey. However, we have no such optically selected group sample for which robust mass estimates are available against which to compare our mass proxies. We therefore turn to simulations.

Applying our mass proxies to a set of groups and clusters generated from simulations, we have the advantage of a known halo mass that is *independent* of any X-ray selection biases. An additional advantage is the availability of a larger sample size. On the other hand, conclusions drawn from simulated systems may be strongly affected by the assumptions and tuning which are incorporated into the simulated galaxy/group population.

For this study, we make use of the mock galaxy catalogues generated for use by the GAMA consortium. The Galaxy And Mass Assembly (GAMA; Driver et al. 2011; Robotham et al. 2011; Merson et al. 2013) project aims to use a broad range of multiwavelength observations to study cosmology and the formation and evolution of galaxies. The optical component of this project is a medium-deep galaxy redshift survey conducted by AAOmega multi-object spectrograph at the Anglo-Australian Observatory. Covering $\sim$250 deg^2 of the sky, the GAMA project has spectroscopy for over 300 000 galaxies to $m_r < 19.8$.

An FoF analysis was conducted by Robotham et al. (2011) to identify galaxy groups within the survey. To support this analysis, a series of mock light cones (Merson et al. 2013) were generated using the Millennium simulation (Springel et al. 2005) populated with a semi-analytic galaxy formation model (Bower et al. 2006). The galaxy LFs of the mocks were adjusted to precisely mimic the observed r-band redshift-dependent GAMA galaxy LF of Loveday et al. (2012). These mocks cover $\sim$144 deg^2 in three 4×12 deg^2 regions, mimicking the original GAMA-I equatorial fields (Driver et al. 2011).

From these light cones, we select group galaxies using the methodology described in Section 2.2.1. We cut the initial galaxy catalogue to $m_r \leq 17.5$ to match our observational sample. Our volumetric group sample is then extracted from this catalogue, centring each extracted cylinder at the position of the original Robotham et al. (2011) GAMA FoF groups containing 10 or more galaxies. In addition, we run our implementation of the E04 FoF algorithm on the mock galaxy catalogue to construct an FoF group sample, retaining only groups with at least 10 members. We define the dominant group halo as the dark matter halo that contributes the most galaxies to either the initial GAMA FoF group (in the case of the volumetric sample) or to our FoF group for the FoF group sample. The mass of this halo is taken to be the true mass of the group. These masses, originally M_{Dhalo} (Jiang et al. 2014), have been converted to our cosmology and to M_{500} using a mass-dependent scaling of $M_{500} = 10^{0.34} M_{\mathrm{Dhalo}}^{0.96}$. This conversion was derived from a set of median halo masses at $z = 0$ for the Millennium-I cosmology (Jiang et al. 2014). To be comparable to our observational sample, we restrict the redshift range of the mock sample to $0.01 \leq z \leq 0.1$ and only consider groups with $M_{500} \geq 10^{13}\,\mathrm{M}_{\odot}$. The final simulated volumetric and FoF group samples consist of 179 and 313 groups, respectively, across 9 mock realisations, a significant improvement in sample size relative to our observational sample.

The larger number of FoF groups, compared to volumetric ones, results from the fact that our FoF algorithm is more generous in linking galaxies than that of Robotham et al. (2011), as discussed below. These mock groups have halo M_{500} ranging from $\sim 10^{13}$ to $\sim 4 \times 10^{14}\,\mathrm{M}_{\odot}$. This upper mass is considerably lower than the highest mass cluster in the observational sample (Abell 2142 with $M_{500} \sim 1.3 \times 10^{15}\,\mathrm{M}_{\odot}$) due to the larger volume probed by the X-ray + SDSS sample, compared to the GAMA mocks.

Considering briefly the performance of our FoF algorithm, it is clear that it recovers a greater number of FoF groups than the GAMA algorithm. As we discussed in Section 2.2.1, our algorithm is essentially that developed by Eke et al. (2004) to generate the 2PIGG group catalogue from the 2dF galaxy redshift survey. Conversely, the GAMA FoF algorithm has been carefully tuned on the GAMA mocks to optimize the fidelity of the grouping. The performance of the 2PIGG algorithm was examined by Robotham et al. (2010) finding that the 2PIGG algorithm was generous in terms of linking together subgroups. By application to the GAMA mocks, we similarly find that the 2PIGG algorithm generously links

Table 5. BCES regression results for the mock group sample corresponding to the best-fitting regression lines in Fig. 8.

Method	α	β	$\log_{10}(x_0)$	ρ^a	$\sigma_{\log M,\,\mathrm{tot}}{}^b$	$\sigma_{\log M,\,\mathrm{int}}{}^{b,c}$	$\sigma_{\log M,\,\mathrm{cal}}{}^b$	$\sigma_{\log M,\,\mathrm{sys}}{}^{b,d}$
				Volumetric selection				
Richness	$1.49^{+0.10}_{-0.11}$	$-0.10^{+0.02}_{-0.02}$	2.10	0.79	0.23	$0.19^{+0.03}_{-0.02}$	$0.03^{+0.00}_{-0.00}$	$0.19^{+0.03}_{-0.02}$
Luminosity	$1.41^{+0.13}_{-0.16}$	$-0.06^{+0.02}_{-0.03}$	11.80	0.68	0.34	$0.29^{+0.04}_{-0.04}$	$0.05^{+0.01}_{-0.01}$	$0.30^{+0.04}_{-0.04}$
Overdensity (N)	$0.87^{+0.08}_{-0.10}$	$0.01^{+0.03}_{-0.03}$	14.27	0.67	0.36	$0.31^{+0.04}_{-0.03}$	$0.05^{+0.01}_{-0.01}$	$0.32^{+0.04}_{-0.03}$
Overdensity (L)	$0.82^{+0.09}_{-0.11}$	$0.03^{+0.03}_{-0.03}$	14.31	0.62	0.42	$0.38^{+0.05}_{-0.04}$	$0.07^{+0.01}_{-0.01}$	$0.38^{+0.05}_{-0.04}$
Dynamical ($r_{\mathrm{rms}}\sigma^2$)	$1.24^{+0.11}_{-0.14}$	$-0.01^{+0.03}_{-0.03}$	8.14	0.62	0.37	$0.34^{+0.05}_{-0.05}$	$0.05^{+0.01}_{-0.01}$	$0.34^{+0.05}_{-0.05}$
Dynamical (σ^3)	$0.87^{+0.07}_{-0.08}$	$0.05^{+0.03}_{-0.03}$	8.23	0.64	0.34	$0.31^{+0.04}_{-0.04}$	$0.05^{+0.01}_{-0.01}$	$0.32^{+0.04}_{-0.04}$
Radius	25.40^{+970}_{-968}	$0.01^{+8.60}_{-9.55}$	2.70	0.07 (66%)	2.25	$2.08^{+66.5}_{-78.2}$	1.28^{+38700}_{-41900}	2.44^{+38700}_{-41900}
				FoF selection				
Richness	$1.35^{+0.07}_{-0.07}$	$-0.17^{+0.02}_{-0.02}$	2.17	0.70	0.31	$0.29^{+0.03}_{-0.02}$	$0.03^{+0.00}_{-0.00}$	$0.29^{+0.03}_{-0.02}$
Luminosity	$1.28^{+0.08}_{-0.09}$	$-0.13^{+0.02}_{-0.03}$	11.86	0.64	0.41	$0.38^{+0.04}_{-0.04}$	$0.04^{+0.01}_{-0.01}$	$0.39^{+0.04}_{-0.04}$
Overdensity (N)	$0.76^{+0.05}_{-0.06}$	$-0.04^{+0.03}_{-0.03}$	14.14	0.66	0.45	$0.42^{+0.04}_{-0.04}$	$0.05^{+0.01}_{-0.01}$	$0.43^{+0.04}_{-0.04}$
Overdensity (L)	$0.71^{+0.05}_{-0.06}$	$0.00^{+0.03}_{-0.04}$	14.18	0.63	0.52	$0.50^{+0.07}_{-0.06}$	$0.05^{+0.01}_{-0.01}$	$0.50^{+0.07}_{-0.07}$
Dynamical ($r_{\mathrm{rms}}\sigma^2$)	$1.08^{+0.06}_{-0.07}$	$-0.10^{+0.02}_{-0.02}$	8.17	0.66	0.36	$0.34^{+0.03}_{-0.03}$	$0.04^{+0.00}_{-0.00}$	$0.34^{+0.03}_{-0.03}$
Dynamical (σ^3)	$0.89^{+0.05}_{-0.06}$	$-0.07^{+0.02}_{-0.02}$	8.00	0.68	0.35	$0.33^{+0.03}_{-0.02}$	$0.03^{+0.00}_{-0.00}$	$0.33^{+0.03}_{-0.02}$
Radius	$5.90^{+0.94}_{-1.41}$	$-0.13^{+0.06}_{-0.08}$	2.86	0.27	1.01	$0.98^{+0.16}_{-0.24}$	$0.19^{+0.07}_{-0.11}$	$1.00^{+0.17}_{-0.25}$

Notes. [a]Spearman rank correlation. Unless shown in parentheses, all results have >95 per cent significance.
[b]Estimated fractional scatters in mass in dex.
[c]Intrinsic scatter in mass accounting for statistical scatter in the proxy (see equation 17). Statistical fractional scatter in mass and proxy are assumed to be equivalent to the mean fractional error from the observational sample reported in Table 3.
[d]The systematic scatter in mass, $\sigma^2_{\log M,\mathrm{sys}} = \sigma^2_{\log M,\mathrm{int}} + \sigma^2_{\log M,\mathrm{int}}$.

together neighbouring structures. Indeed, 18 per cent of groups have more than 50 per cent contamination (i.e. of the linked members, less than half belong to the dominant halo), whereas this figure is only 10 per cent for GAMA FoF groups. As the 2PIGG algorithm is more generous, we find it recovers a high fraction of true members, where 64 per cent of our groups link together more than 90 per cent of the dominant halo's galaxies, compared to only 10 per cent of the GAMA FoF groups.

In summary, our algorithm will link most of the true member galaxies, but will also include a larger interloper fraction. Given that the GAMA galaxy LF at low redshift is similar to that of SDSS (Loveday et al. 2012), it is likely that these conclusions are also applicable to the use of our algorithm on our SDSS-based observational sample.

In the light of these differences in FoF performance, we have also examined the performance of our mass proxies on the original GAMA mock FoF catalogue (cut to our mass, magnitude and redshift range). We find little significant difference to the results with our own FoF groups, the most notable difference being that the richness overdensity and σ^3 proxies perform closer to theoretical expectations. For consistency with the observational sample, we proceed using our own FoF selection, with reference to the original GAMA FoF selection where differences are significant.

7 COMPARING RESULTS FOR OBSERVED AND MOCK GROUPS

In this section, we discuss the results from the mock groups. We calibrate each mass–proxy relation, again using BCES (X|Y) regression, with the intention of examining whether they are consistent with observations. We do not attempt to bootstrap proxy errors for the mock groups. Instead, as we have selected mock groups to be comparable to our observational sample, we approximate the fractional statistical error in proxy and mass for all mock groups as the root-mean-square fractional statistical error in mass and proxy from the observational sample. Our analysis here is otherwise identical to that of the observational sample in Section 5.

The figures in this section follow those of Section 5, except that, for comparison to the observational sample, we overplot the mock mass–proxy data with the observational calibration line and error region. To aid in readability and highlight any trend, we additionally bin our mock group's data. To avoid any biases introduced by the richness cut in the original GAMA catalogue this binning is performed horizontally, in $\log_{10}(M_{500})$ slices.

7.1 Richness and luminosity

With Table 5 and the top row of Fig. 8 we find that these GAMA mocks produce a mass–aperture richness relation that is in good agreement with observations, whilst the mass–aperture luminosity relation is shallower, but marginally consistent within the lower bound of the observational result. In general terms, these slopes are also consistent with the effect of an aperture, though not both an aperture and SFE variation. Conversely, the FoF richness and luminosity estimators give results discrepant with the observational sample, both being significantly steeper.

We again consider the richness–luminosity relations, Fig. 9. We find that the relation, whilst in good agreement with the observations and expected slope of 1 at high mass, deviates from the power-law slope at low mass. Specifically, we see an underluminous population of mock groups at low mass. This is possibly related to the galaxy formation models' implementation of satellite galaxy merging where a satellite galaxy is not resolved in its subhalo (Robotham et al. 2011). If assigned too long a merging time, this will result in an excess of faint satellite galaxies in close orbits of central galaxies or the centre of the halo. This close satellite excess will also affect other mass proxies.

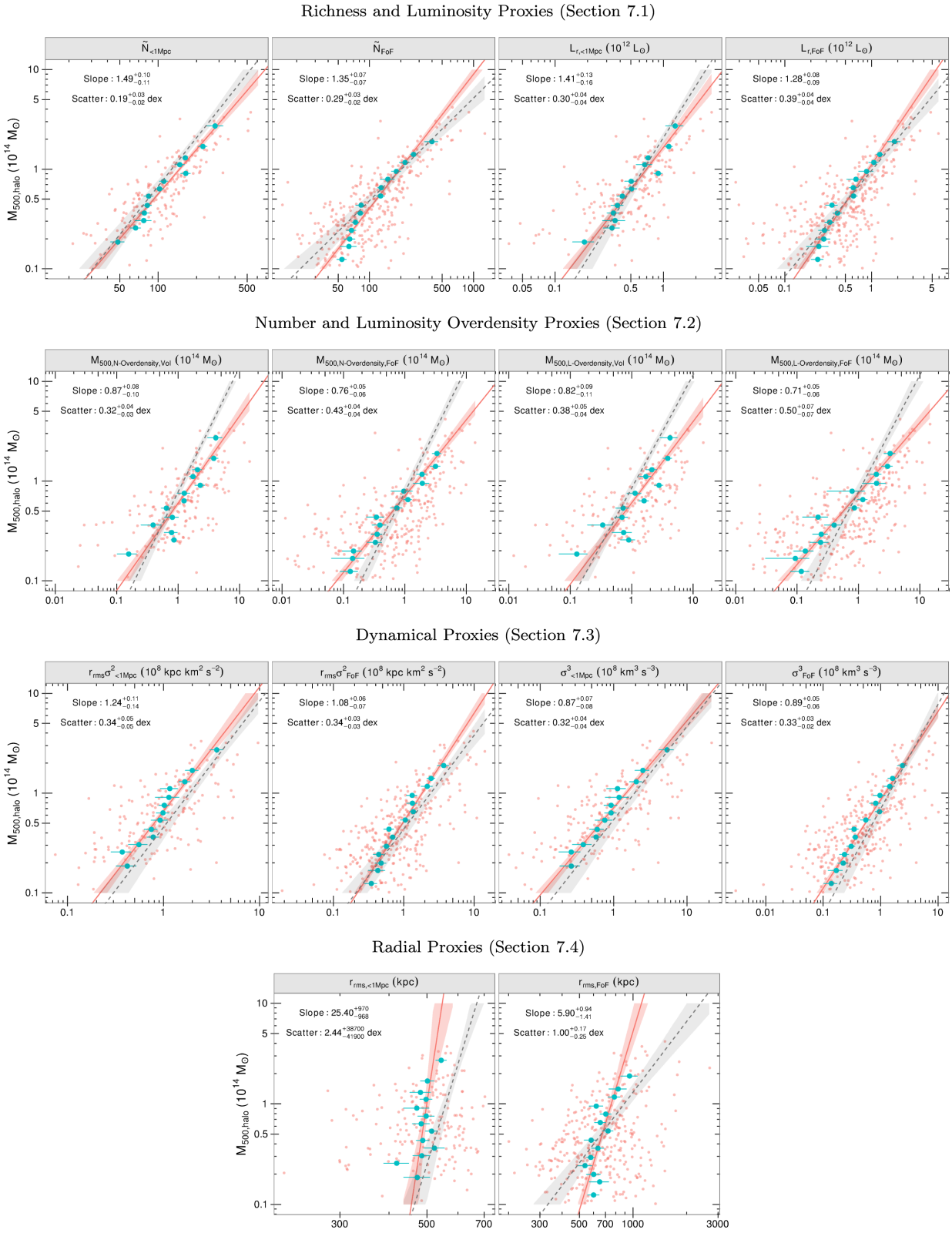

Figure 8. Mock group halo M_{500} plotted against each mass proxy with, from left to right, top to bottom, Richness, Luminosity, Number Overdensity, Luminosity Overdensity, $r_{\rm rms}\sigma^2$ Dynamical, σ^3 Dynamical and $r_{\rm rms}$ radius. Alternating plots show results for the volumetric and FoF sample. Faint red points show all mock groups whilst the blue points show horizontally binned averages and their standard errors. The red solid lines and shaded regions show the BCES(X|Y) regressions and a 68 per cent interval drawn from the distribution of bootstrapped fit results for the mock data. The grey dashed lines and shaded regions show the same for the observational sample shown in Fig. 2.

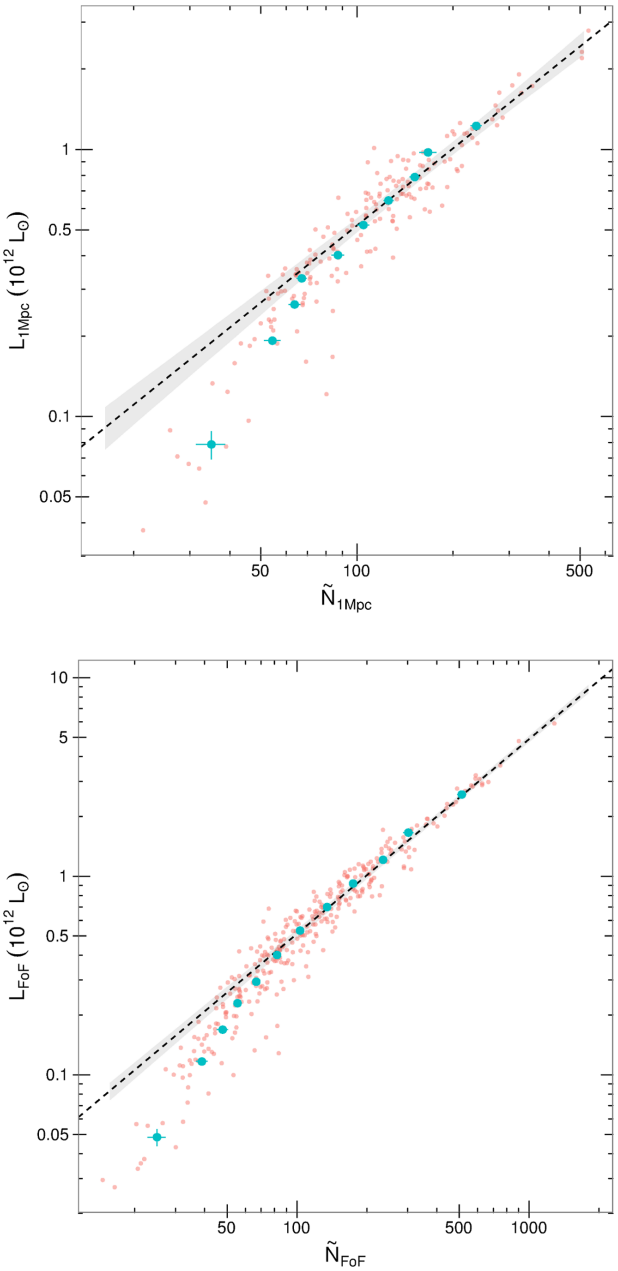

Figure 9. As Fig. 8 with top: background-subtracted optical luminosity against richness within 1 Mpc, and bottom: total FoF luminosity against FoF richness.

7.2 Overdensity

As shown in Table 5 and the second row of Fig. 8, we find that the mock groups have a significantly shallower relation for overdensity masses than that of both the observed groups and the expectation of unity. This is true regardless of the selection method. In the case of the observed groups, we concluded that the steeper than unity slopes could be understood in the light of expected variations in SFE with halo mass. Another factor here may be the previously mentioned anomalies in the radial distribution of galaxies within mock groups (Robotham et al. 2011; Han et al. 2015).

For comparison, the GAMA FoF selection gives a slope of $0.95^{+0.12}_{-0.15}$ for the richness overdensity proxy, consistent with unity. Luminosity overdensity is also steeper for the GAMA FoF sample than for ours, though not significantly, and is not consistent with unity.

7.3 Dynamical

The results of the $r_{\rm rms}\sigma^2$ and σ^3 proxies are also shown in Table 5 and in the third row of Fig. 8. As with the observational sample, we find that it is difficult to draw conclusions from the volumetric $r_{\rm rms}\sigma^2$ relation due to the steepness of the $r_{\rm rms}$ relation (see below). Interestingly, we find that in contrast to the situation with the observed groups, the two σ^3 proxies, are also significantly discrepant from the expectation of unity slope. For comparison, the GAMA FoF selection gives a slope for the σ^3 proxy of $0.97^{+0.09}_{-0.09}$, consistent with unity.

7.4 Radii

Table 5 and the final row of Fig. 8 show the results of the $r_{\rm rms}$ estimator. We first observe that the total scatter on this proxy is substantial for both selection methods (2.3 and 1.0 dex for the volumetric and FoF samples, respectively). This makes it difficult to reliably constrain any calibration relation. Despite this, we find that the volumetric sample, which is again subject to interloper contamination and aperture effects, gives a relation that is steeper than the already very steep observed relation. The FoF estimator is also significantly steeper than the observational relation, and also steeper than the expected slope of 3.

7.5 Summary

In Table 6, we bring together a summary of the comparison between the mass–proxy relations derived from the observational sample and the mocks, and a comparison of these with theoretical expectations. The proxies based on the mocks have relations which differ significantly from observations, with the exception of the σ^3 dynamical estimator, which agrees within the errors on the observational result. We also note that whilst the radius–mass relation is significantly different to the observationally derived relation, due to known issues with the orbits of satellite galaxies within the mocks (Robotham et al. 2011; Han et al. 2015), it behaves in a similar fashion, with an extremely steep relation for the volumetric sample.

The discrepancy for luminosity and overdensity might be explained by a difference in LF between that assumed by the mocks and the observations. The semi-analytic galaxies were adjusted to reproduce a global LF that matches that of the real GAMA fields (Robotham et al. 2011); however, it does not follow that they will reproduce the LF within groups and clusters. A study by McNaught-Roberts et al. (2014) examining the LF of different density environments within the real GAMA fields and the GAMA mocks revealed no significant difference between simulations and reality. We note however that this study looked at the local density of each galaxy, rather than classifying galaxies as belonging to groups or the field.

The small difference between the global LFs measured for the SDSS (Blanton et al. 2003) and GAMA (Loveday et al. 2012) is unlikely to contribute greatly to the unexpectedly flat slope of the $M_{\rm overdensity}$–$M_{\rm halo}$ relations for the mocks. A more likely cause is the radial distribution of galaxies in the mocks. As discussed in Han et al. (2015, see also Robotham et al. 2011), who compared mass-observable relations for the GAMA fields and GAMA mocks, there is tension between the mocks and real data for the relationship between mass and r_{50} (the radius containing 50 per cent of a group's

Table 6. Summary – comparison of the performance of the observations and mock results, noting any deviation from theoretical expectations.

Method	Observational	Mocks
	Volumetric selection	
Richness	Steeper than naive expectation of unity. Plausibly modified by aperture and SFE. Low systematic scatter.	Slope steeper than unity. Plausibly modified by aperture effects. Rough agreement with observational sample.
Luminosity	Steeper than naive expectation of unity. Plausibly modified by aperture and SFE.	Slope steeper than unity. Plausibly modified by aperture effects. Rough agreement with observational sample.
Overdensity (N)	Steeper than expected slope of unity. Plausibly modified by SFE and mass-to-light. Low systematic scatter, but high statistical scatter.	Shallower than expected slope of unity.
Overdensity (L)	Steeper than expected slope of unity. Plausibly modified by SFE and mass-to-light. Low systematic scatter, but highest statistical scatter.	Shallower than expected slope of unity.
Dynamical ($r_{\rm rms}\sigma^2$)	Steeper than expected slope of unity. Possibly unreliable due to poor quality $r_{\rm rms}$.	Steeper than expected slope of unity. Possibly unreliable due to poor-quality $r_{\rm rms}$.
Dynamical (σ^3)	Consistent with expected slope of unity.	Shallower than expected slope of unity, but consistent within the errors on the observational sample.
Radius	Slope much steeper than naive expectation of 3. Plausibly modified by aperture effects. Largest systematic scatter.	Slope much steeper than naive expectation of 3. Plausibly modified by aperture effects. Large scatter observed in the data.
	FoF selection	
Richness	Consistent with expectations of unity before and after accounting for SFE. Low systematic scatter.	Steeper than expected slope of unity.
Luminosity	Consistent with expectations of unity before and after accounting for SFE.	Steeper than expected slope of unity.
Overdensity (N)	Steeper than expected slope of unity. Plausibly modified by SFE and mass-to-light. Low systematic scatter, but high statistical scatter.	Shallower than expected slope of unity.
Overdensity (L)	Steeper than expected slope of unity. Plausibly modified by SFE and mass-to-light. Low systematic scatter, but highest statistical scatter.	Shallower than expected slope of unity.
Dynamical ($r_{\rm rms}\sigma^2$)	Consistent with expected slope of unity.	Steeper than expected slope of unity.
Dynamical (σ^3)	Consistent with expected slope of unity.	Shallower than expected slope of unity. Consistent within the errors on the observational sample.
Radius	Shallower than expected slope of 3. Large systematic scatter.	Significantly steeper than expected slope of 3.

projected galaxies) which may have its origin in the approximate treatment of dynamical friction within the semi-analytic model, and consequent inaccuracies in the rate of orbital decay. This can be expected to modify the relationship with halo mass for both the overdensity and radius proxies.

Han et al. (2015) also find a systematic bias in the mass–velocity dispersion relation for the mocks compared to the real groups, again, due to known issues with the small-scale dynamics of the GAMA mocks (Robotham et al. 2011). This may be the cause of the lower slope in the mock mass–σ^3 relation relative to observation and to theoretical expectations.

Comparing the residuals of predicted masses with respect to the halo mass, in a similar fashion to that discussed in Section 5.6, we find much weaker correlation between unrelated proxies than was apparent in the observational sample. For example, the richness and σ^3 estimators show a weak, but still significant (>95 per cent) correlation ($\rho = 0.18^{+0.08}_{-0.08}$ and $0.31^{+0.06}_{-0.05}$ for volumetric and FoF samples, respectively) significantly different to the values of $0.71^{+0.16}_{-0.11}$ and $0.66^{+0.19}_{-0.15}$ seen observationally. The asymmetry in Fig. 10 can be compared with that in the corresponding observational plot, Fig. 7. The small error range from the mocks' bootstrapped correlations agrees with the high level of significance indicating that the mocks' residuals behave very differently to those of the observational sample, supporting the possibility that this effect is due to excess X-ray mass errors. The non-zero correlation seen for the mocks may arise from the effects of interlopers enhancing both richness and velocity dispersion. The fact that it is stronger in the FoF case, where no correction of richness for background galaxies has been included, supports this hypothesis.

Comparing Table 5 with Table 3, it can be seen that the total scatter, $\sigma_{\log M,\,\rm tot}$, about the mass–proxy relations for the mocks are mostly greater than the corresponding scatter for the observational sample. This suggests that, as expected, we are sampling a more diverse group population in the mocks, in terms of their dynamics and evolutionary state, compared to the relaxed X-ray-bright systems on which our observational sample is based. This emphasizes the importance of allowing for selection effects in any study of galaxy groups – samples selected in different ways can

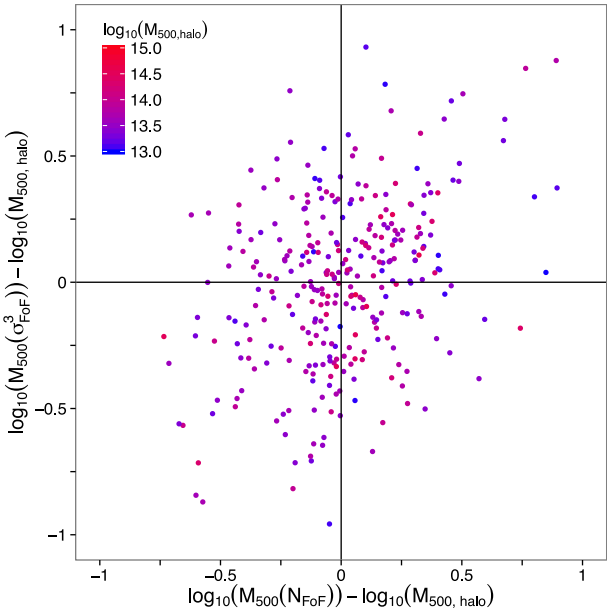

Figure 10. Fractional residuals about the regression lines for the mock groups' FoF richness and σ^3 dynamical estimators defined as $R = \log_{10}(M_{500}(x)) - \log_{10}(M_{500,\,\mathrm{halo}})$. Each data point is coloured according to its halo mass, with blue $\sim 10^{13}\,\mathrm{M}_\odot$ and red $\sim 10^{15}\,\mathrm{M}_\odot$.

be expected to catch a different mix of dynamical and evolutionary states.

Finally, we comment briefly on the use of these proxies when the true galaxy membership is known. Thus far we have analysed the mocks by applying similar selections processes (volumetric and FoF) to those used on the observational data. Since our mass–proxy relations have been calibrated on observational data which will inevitably be contaminated by interlopers, we can expect to find different results for mocks groups if we use the *true* group membership, which is available for these simulated systems. Examining mass–proxy relations using the true galaxy membership can cast light on the fundamental nature of these proxies. Specifically considering the mass–richness relation (scaled to $M_r < -16.5$) and the mass–σ^3 these fit with BCES (X|Y) slopes of 1.08 ± 0.03 and 0.81 ± 0.05 with systematic scatters of 0.13 ± 0.03 dex and 0.29 ± 0.03 dex, respectively, again assuming the masses and proxies have statistical errors comparable to the mean errors from the observational sample. Interestingly, the first relation, whilst significantly shallower than the mock mass–richness relations derived earlier, is still steeper than the expectation of unity by more than 2σ, even in the absence of aperture effects or FoF selection effects. This is likely related to the way galaxies populate the haloes within the mocks, and to their merger history. Conversely, the mass–σ^3 relation is shallower than unity, as with other mock mass–σ^3 relations, illustrating the known concerns with the dynamics of the mock haloes (Robotham et al. 2011; Han et al. 2015). Both relations show smaller systematic scatter than the corresponding relations with volumetric of FoF selection, indicating the scattering effects of interlopers.

8 DISCUSSION AND CONCLUSIONS

In this paper, we have investigated the behaviour of mass proxies based upon richness, luminosity, galaxy number and luminosity overdensity, characteristic radii and dynamics. We have applied these estimators to sets of group galaxies derived in two ways: a σ-clipped cylindrical volume and an FoF analysis. The mass proxies have been calibrated using masses derived from the X-ray properties of the hot intragroup gas, and the resulting calibrations compared with theoretical expectations. The performance of the same mass proxies has been assessed on a set of mock groups for which true halo masses are known. Our findings for each proxy are summarized in Table 6. We conclude that:

(i) The calibrated relationships between these mass proxies and the X-ray-derived masses are consistent with theoretical expectations once aperture effects and the impact of mass-dependent variations in SFE are allowed for.

(ii) The most reliable mass estimate is provided by modelling the galaxy overdensity (number or luminosity) within a 5 Mpc cylinder (i.e. for a volumetric group sample). These mass proxies exhibit the smallest intrinsic scatter and the least uncertainty introduced by the calibration, resulting in only 0.14–0.15 dex (≈33 per cent) systematic uncertainty in any final mass estimate. However, for our data set these proxies also have the largest statistical scatter, and are therefore only likely to work well where one has high-quality galaxy membership data, such that reliable galaxy density profiles can be fitted.

(iii) Richness, for either volumetric or FoF selection, is the next best method, and has the merit of being simple to apply, and working well even for fairly small galaxy samples. Of course, it is essential that this, and the overdensity and luminosity proxies, should be corrected for any incompleteness in survey coverage or spectroscopic observations.

(iv) Dynamical estimators show larger scatter, with a systematic uncertainty of ≈0.3 dex (≈70 per cent). However, they do have the attractive feature that they are relatively robust against spectroscopic incompleteness. They may therefore become the preferred option for mass estimation where incompleteness is difficult to quantify. The $r_{\mathrm{rms}}\sigma^2$ proxy should be avoided, in favour of σ^3, for volumetric galaxy samples, due to bias in the estimation of r_{rms}.

(v) An r_{rms} radial estimator should not be used in cases where a restrictive aperture has been applied, such as our volumetric selection. It is more useful in the case of an FoF sample; however, the scatter is substantial (0.35 dex, or 81 per cent, systematic uncertainty).

The same set of mass proxies was applied to groups drawn from the mock galaxy data generated for the GAMA project. We find that in most cases the mass–proxy relations differ from those for the observational group sample, and that the scatter of the sample around the mean relationship is greater. This larger scatter suggests that our mock sample spans a more diverse collection of galaxy groups than the relaxed X-ray-bright systems which constitute the observational sample.

This greater diversity in the population might account for some of the differences in the mass–proxy relations which we see. However, the discrepancies between observational and mock samples are most striking for the overdensity and radius–mass proxies, and these are affected by the radial distribution of group galaxies. This distribution is unreliable in the inner regions of the mock groups, due to the approximate treatment of orbital decay in the semi-analytic model on which they are based (Robotham et al. 2011; Han et al. 2015). The richness, luminosity and σ^3 mass proxies are in better agreement for the observational and mock samples, though in general the fitted relations do not agree within the 1σ range.

The estimated systematic error ($\sigma_{\log M,\,\mathrm{sys}}$) for each mass proxy, shown in the last column of Table 3, is an important quantity. This represents the uncertainty in the $\log_{10}$(mass) which results for each

mass proxy even when the value of the proxy is precisely known. The low values of this for the overdensity proxies when applied to the volumetric sample are the basis for our statement above that this provides the most reliable single mass estimator where high-quality galaxy data are available.

However, our estimate of this systematic error is subject to two opposing biases. As explained in Section 4.2, this quantity is derived from the observed scatter in mass about the fitted mass–proxy relation by subtracting the estimated statistical contributions to this scatter and then adding the variance which arises from uncertainty in the calibrated power-law relation. However, we have seen (Section 5.6) that for a subset of our observational sample there appear to be errors in the X-ray-derived masses which exceed their statistical error estimates. Since we have not allowed for these, this will have inflated our estimates of $\sigma_{\log M,\,\mathrm{sys}}$. On the other hand, we may expect the scatter in intrinsic group properties for a group sample more diverse than our observational sample to be larger, and this would naturally lead to *larger* scatter in the relation between true mass and proxy-mass estimates, as we see from our mock sample. On the whole, it is probably best to regard the systematic error estimates for the mass proxies in Table 3 as being a useful indication, which is likely to be pessimistic for use with relaxed group samples, but not necessarily when used on a more diverse set of groups.

The diversity in the properties of galaxy groups has been documented by many authors (e.g. Zabludoff & Mulchaey 1998; Osmond & Ponman 2004; Rasmussen et al. 2006; Balogh et al. 2011). In the light of this, there are considerable advantages to employing not just a single mass proxy, but a whole set of them, since different proxies will be robust against different effects. For example, mass estimates based on velocity dispersion are likely to be in error for dynamically disturbed groups which have recently experienced a major merger, whilst an estimate based on galaxy richness or luminosity should be relatively unaffected. On the other hand, variations in SFE, or in the fraction of stars stripped from galaxies will impact on richness, luminosity and overdensity estimates, but not on the velocity dispersion or radius proxies. The effects of orbital decay and merging in old groups will impact on the richness and the fitted galaxy profile, though the impact on the total overdensity within r_{500} should be modest. It would be very interesting to compare the performance of these different mass proxies across a wide range of groups and clusters with high-quality X-ray data and optical photometry to explore such effects. This would also highlight any differences in behaviour in our mass proxies between X-ray-selected and optically selected group and cluster samples.

Finally, we remind the reader that the proxy–mass calibrations presented here are derived from a low-redshift sample of relaxed groups and clusters, and should not be used without modification for systems at significant ($z \gg 0.1$) redshifts. In some cases, the nature of the required modification can be simply predicted theoretically – for example, the velocity dispersion of a virialized system of given mass should scale as $H(z)^{1/3}$, where $H(z)$ is the Hubble parameter at redshift z – but in others (the effects of galaxy luminosity evolution, for example) calibration on a high-redshift sample is really required.

ACKNOWLEDGEMENTS

RJP acknowledges the support from a Science and Technology Facilities Council Postgraduate Studentship, and also thanks Sean McGee and Scott Kay for their insightful discussion of this work during thesis examination. PN acknowledges the support of the Royal Society through the award of a University Research Fellowship and the European Research Council, through receipt of a Starting Grant (DEGAS-259586).

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Ahn C. P. et al., 2014, ApJS, 211, 17
Akritas M. G., Bershady M. A., 1996, ApJ, 470, 706
Alard C., 2013, MNRAS, 428, 340
Andreon S., Congdon P., 2014, A&A, 568, A23
Andreon S., Hurn M. A., 2010, MNRAS, 404, 1922
Arnaud M., Evrard A. E., 1999, MNRAS, 305, 631
Balogh M. L., Mazzotta P., Bower R. G., Eke V., Bourdin H., Lu T., Theuns T., 2011, MNRAS, 412, 947
Bartelmann M., 1996, A&A, 313, 697
Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Berlind A. A. et al., 2006, ApJS, 167, 1
Biviano A., Murante G., Borgani S., Diaferio A., Dolag K., Girardi M., 2006, A&A, 456, 23
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Blanton M. R. et al., 2003, ApJ, 592, 819
Blanton M. R. et al., 2005, AJ, 129, 2562
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Budzynski J. M., Koposov S. E., McCarthy I. G., McGee S. L., Belokurov V., 2012, MNRAS, 423, 104
Budzynski J. M., Koposov S. E., McCarthy I. G., Belokurov V., 2014, MNRAS, 437, 1362
Carlberg R. G. et al., 1997, ApJ, 485, L13
Colless M. et al., 2001, MNRAS, 328, 1039
Davis M., Tonry J., Huchra J., Latham D. W., 1980, ApJ, 238, L113
Driver S. P. et al., 2011, MNRAS, 413, 971
Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, MNRAS, 390, L64
Eke V. R. et al., 2004, MNRAS, 348, 866 (E04)
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Gonzalez A. H., Zabludoff A. I., Zaritsky D., 2005, ApJ, 618, 195
Han J. et al., 2015, MNRAS, 446, 1356
Hansen S. M., McKay T. A., Wechsler R. H., Annis J., Sheldon E. S., Kimball A., 2005, ApJ, 633, 122
Hogg D. W., Bovy J., Lang D., 2010, preprint (arXiv:1008.4686)
Ikebe Y., Reiprich T. H., Böhringer H., Tanaka Y., Kitayama T., 2002, A&A, 383, 773
Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, MNRAS, 440, 2115
Kaiser N., 1986, MNRAS, 222, 323
Kelly B. C., 2007, ApJ, 665, 1489
Leauthaud A. et al., 2012, ApJ, 746, 95
Loveday J. et al., 2012, MNRAS, 420, 1239
McNaught-Roberts T. et al., 2014, MNRAS, 445, 2125
Mamon G. A., Biviano A., Murante G., 2010, A&A, 520, A30
Merloni A. et al., 2012, preprint (arXiv:1209.3114)
Merson A. I. et al., 2013, MNRAS, 429, 556
Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Mulchaey J. S., 2000, ARA&A, 38, 289
Narayan R., Bartelmann M., 1996, preprint (arXiv:astro-ph/9606001)
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neumann D. M., Arnaud M., 2001, A&A, 373, L33
Osmond J. P. F., Ponman T. J., 2004, MNRAS, 350, 1511
Ostriker J. P., Peebles P. J. E., Yahil A., 1974, ApJ, 193, L1
Padmanabhan N. et al., 2008, ApJ, 674, 1217
Panter B., Jimenez R., Heavens A. F., Charlot S., 2007, MNRAS, 378, 1550
Parker L. C., Hudson M. J., Carlberg R. G., Hoekstra H., 2005, ApJ, 634, 806
Ponman T. J., Cannon D. B., Navarro J. F., 1999, Nature, 397, 135
Popesso P., Böhringer H., Romaniello M., Voges W., 2005, A&A, 433, 415
Popesso P., Biviano A., Böhringer H., Romaniello M., 2007, A&A, 464, 451

R Development Core Team 2009, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna
Ramella M., Boschin W., Geller M. J., Mahdavi A., Rines K., 2004, AJ, 128, 2022
Rasmussen J., Ponman T. J., Mulchaey J. S., Miles T. A., Raychaudhury S., 2006, MNRAS, 373, 653
Robotham A., Wallace C., Phillipps S., De Propris R., 2006, ApJ, 652, 1077
Robotham A., Phillipps S., de Propris R., 2010, MNRAS, 403, 1812
Robotham A. S. G. et al., 2011, MNRAS, 1172
Rozo E. et al., 2009, ApJ, 703, 601
Rozo E., Rykoff E., Koester B., Nord B., Wu H.-Y., Evrard A., Wechsler R., 2011, ApJ, 740, 53
Rykoff E. S. et al., 2012, ApJ, 746, 178
Sanderson A. J. R., Ponman T. J., 2010, MNRAS, 402, 65 (SP10)
Schechter P., 1976, ApJ, 203, 297
Springel V. et al., 2005, Nature, 435, 629
Sun M., Voit G. M., Donahue M., Jones C., Forman W., Vikhlinin A., 2009, ApJ, 693, 1142 (S09)
Voges W. et al., 1999, A&A, 349, 389
Voit G. M., Bryan G. L., 2001, Nature, 414, 425
Zabludoff A. I., Mulchaey J. S., 1998, ApJ, 496, 39
Zandivarez A., Martínez H. J., 2011, MNRAS, 415, 2553
Zibetti S., White S. D. M., Schneider D. P., Brinkmann J., 2005, MNRAS, 358, 949
Zwicky F., 1937, ApJ, 86, 217

APPENDIX A: GENERATIVE MODEL

We begin by assuming that the physics of group formation produces physical masses and proxies that follow a (correlated) two-dimensional lognormal distribution:

$$\log M_{500}^{\rm true}, \log x_{\rm true} \sim N\left(\mu, \Sigma\right) \tag{A1}$$

with

$$\mu = \begin{pmatrix} \mu_{500} \\ \mu_x \end{pmatrix} \tag{A2}$$

and

$$\Sigma = \begin{pmatrix} \Sigma_{500} & \Sigma_{500x} \\ \Sigma_{500x} & \Sigma_x \end{pmatrix}. \tag{A3}$$

We further assume that, conditioned on the true values of $M_{500}^{\rm true}$ and $x_{\rm true}$, the observed masses and proxies follow a lognormal distribution with standard deviation equal to the reported observational errors

$$\log M_{500} \sim N\left(\log M_{500}^{\rm true}, \sigma_{500}\right) \tag{A4}$$

$$\log x \sim N\left(\log x_{\rm true}, \sigma_x\right). \tag{A5}$$

Given an observation consisting of $d = \{\log M_{500}, \log x, \sigma_{500}, \sigma_x\}$, the likelihood of the observation conditional on the parameters μ and Σ can be obtained by integrating over the unobserved $M_{500}^{\rm true}$ and $x_{\rm true}$:

$$p\left(d|\mu, \Sigma\right) = \int {\rm d}M_{500}^{\rm true}\, {\rm d}x_{\rm true}\, p\left(d|M_{500}^{\rm true}, x_{\rm true}\right) \times p\left(M_{500}^{\rm true}, x_{\rm true}|\mu, \Sigma\right) = N\left[\mu, \Sigma'\right]\left(\log M_{500}, \log x\right), \tag{A6}$$

where

$$\Sigma' = \begin{pmatrix} \Sigma_{500} + \sigma_{500}^2 & \Sigma_{500x} \\ \Sigma_{500x} & \Sigma_x + \sigma_x^2 \end{pmatrix}. \tag{A7}$$

Given a training set of observations, $\{d^{(i)}\}$, we can construct the posterior distribution on the parameters μ and Σ implied by these data via Bayes' rule:

$$p\left(\mu, \Sigma|\left\{d^{(i)}\right\}\right) \propto p\left(\mu, \Sigma\right) \prod_i p\left(d^{(i)}|\mu, \Sigma\right), \tag{A8}$$

where we are ignoring normalizing constants that are independent of μ and Σ, and $p(\mu, \Sigma)$ is a prior distribution representing our knowledge of μ and Σ before observing any of the data. In the limit where we have a large training set, the prior becomes irrelevant; here we take the prior to be flat (i.e. constant) in μ and the eigenvalues of Σ. The latter ensures that we are a priori agnostic about the orientation and scale of the error ellipse represented by Σ.

We draw samples from the five-dimensional posterior distribution on μ and Σ using the EMCEE sampler (Foreman-Mackey et al. 2013). Given a posterior on Σ, we can construct a posterior on the directions of the principal axes of the error ellipse in the $M_{500}^{\rm true}$-$x_{\rm true}$ relation fit by our model; the slope of the long-axis direction is the equivalent of a regression parameter in traditional models where $M_{500} = \alpha x + \beta$. We will return to this below, but the best way to use the generative model is not to regard it a mass–proxy calibration relation, but rather to use it to generate a Bayesian posterior probability distribution for the mass, given some proxy measurement.

For a given choice of the parameters μ and Σ, our model predicts the joint distribution of $M_{500}^{\rm true}$, $x_{\rm true}$, M_{500}, and x in future observations:

$$p\left(M_{500}^{\rm true}, x_{\rm true}, M_{500}, x|\mu, \Sigma\right) \propto N\left[M_{500}^{\rm true}, \sigma_{500}\right](M_{500}) \times N\left[x_{\rm true}, \sigma_x\right](x)\, N\left[\mu, \Sigma\right]\left(M_{500}^{\rm true}, x_{\rm true}\right), \tag{A9}$$

where $N[m, \sigma](y)$ denotes a normal distribution for y with mean m and standard deviation σ.

If we have a subsequent observation of only x and σ_x, but not M_{500} or σ_{500}, we can integrate over the unobserved data to obtain the conditional distribution for $M_{500}^{\rm true}$ and $x_{\rm true}$ under the model:

$$\begin{aligned} &p\left(M_{500}^{\rm true}, x_{\rm true}|x, \mu, \Sigma\right) \\ &\propto \int {\rm d}M_{500}\, p\left(M_{500}^{\rm true}, x_{\rm true}, M_{500}|x, \mu, \Sigma\right) \\ &\propto N\left[x_{\rm true}, \sigma_x\right](x)\, N\left[\mu, \Sigma\right]\left(M_{500}^{\rm true}, x_{\rm true}\right). \end{aligned} \tag{A10}$$

Conditional on μ and Σ, this is a normal distribution with

$$\left\langle M_{500}^{\rm true}\right\rangle = \mu_{500} + \frac{(x - \mu_x)\,\Sigma_{500x}}{\sigma_x^2 + \Sigma_x}, \tag{A11}$$

$$\left\langle x_{\rm true}\right\rangle = \frac{\mu_x \sigma_x^2 + x\,\Sigma_x}{\sigma_x^2 + \Sigma_x}, \tag{A12}$$

$${\rm var}\left(M_{500}^{\rm true}\right) = \Sigma_{500} - \frac{\Sigma_{500x}^2}{\Sigma_x + \sigma_x^2}, \tag{A13}$$

and

$${\rm var}\left(x_{\rm true}\right) = \frac{\sigma_x^2 \Sigma_x}{\Sigma_x + \sigma_x^2}. \tag{A14}$$

Note that these equations exhibit several desirable features:

(i) Equations (A11) and (A12) exhibit a regression to the mean. Depending on the measurement error, σ_x, and the degree of correlation between M_{500} and x in the model, the predictions for $M_{500}^{\rm true}$ and $x_{\rm true}$ interpolate between the model averages μ_{500} and μ_x and the measurement and its extrapolation.

(ii) Equation (A13) results in a posterior variance for $M_{500}^{\rm true}$ that is always smaller than the naive variance of $M_{500}^{\rm true}$ under the model. In the limit that the measurement of the proxy is perfect ($\sigma_x \to 0$), the reduction in variance is exactly what would be expected by the correlation coefficient,

$$R^2 \equiv \frac{\Sigma_{500x}^2}{\Sigma_{500}\Sigma_x}. \tag{A15}$$

(iii) Equation (A14) demonstrates that the posterior variance on $x_{\rm true}$ is always smaller than the measurement variance σ_x^2, because of the information encoded in the parameters μ and Σ; the reduction factor is

$$\frac{1}{1+\frac{\sigma_x^2}{\Sigma_x}}. \tag{A16}$$

Given a posterior distribution on μ and Σ from a training data set, we can compute the expected distribution of $M_{500}^{\rm true}$ and $x_{\rm true}$ after a subsequent observation of x with error σ_x using

$$\begin{aligned} p\left(M_{500}^{\rm true}, x_{\rm true}|x, \sigma_x, \left\{d^{(i)}\right\}\right) \\ = \int {\rm d}\mu\, {\rm d}\Sigma\; p\left(M_{500}^{\rm true}, x_{\rm true}|x, \sigma_x, \mu, \Sigma\right) \\ \times p\left(\mu, \Sigma|\left\{d^{(i)}\right\}\right). \end{aligned} \tag{A17}$$

In a similar way, we can compute posterior averages of the quantities in equations (A11)–(A14).

To demonstrate the performance of this generative model in estimating masses, given training set data similar to that available from the present study, we generate simulated mass–proxy data sets following a known relation. For this purpose, we assume the mass–proxy relation given in equation (16) with slope $\alpha = 0.8$ and intercept $\beta = 5$. Each data set consists of 22 groups whose masses are drawn from a log-uniform distribution in the range $13 \lesssim \log_{10}(M/{\rm M}_\odot) \lesssim 15$. We assume that this mass is the primary property of a group and invert equation (16), including intrinsic scatter, to determine a 'proxy' value. Hence, proxy values are generated using

$$\log_{10}(x) = 1/\alpha \log_{10}(M) - \beta/\alpha + N(0, \sigma_{\rm int}),$$

where $N(0, \sigma_{\rm int})$ represents intrinsic Gaussian scatter in the proxy value, and we take $\sigma_{\rm int} = 0.3$ dex.

We also include statistical errors on the data points in our synthetic data sets which are comparable to those on the observed data. To do this, we rank each set of 22 simulated groups by mass, and hence match them to the observed sample. The statistical errors, and corresponding scatter, in mass are then taken from the corresponding observational data. For the errors and scatter on the proxy, we adopt the observational values from the volumetric σ^3 proxy.

We first generate a *training data set* of 22 systems, which is used, as described above, to fit the parameters of the generative model. This model is then used on a further *test data set* of 22 systems, for which we estimate masses based on the measured proxy values and their errors. The posterior probability distributions derived for these 22 test groups are shown, ratioed to their true masses, in Fig. A1, and are found to give unbiased estimates of $M_{500}^{\rm true}$ – that is, over many synthetic observations of the proxy the average difference between the $M_{500}^{\rm true}$ posterior mean and the actual $M_{500}^{\rm true}$ is zero. Note however that, as shown in Table 2, if one were to apply the generative fit as a traditional scaling relation (i.e. not recover masses within the Bayesian framework), one would not recover unbiased masses.

Fig. A2 shows the results of mass prediction using the generative model. Using a test data set, we plot the measured proxy

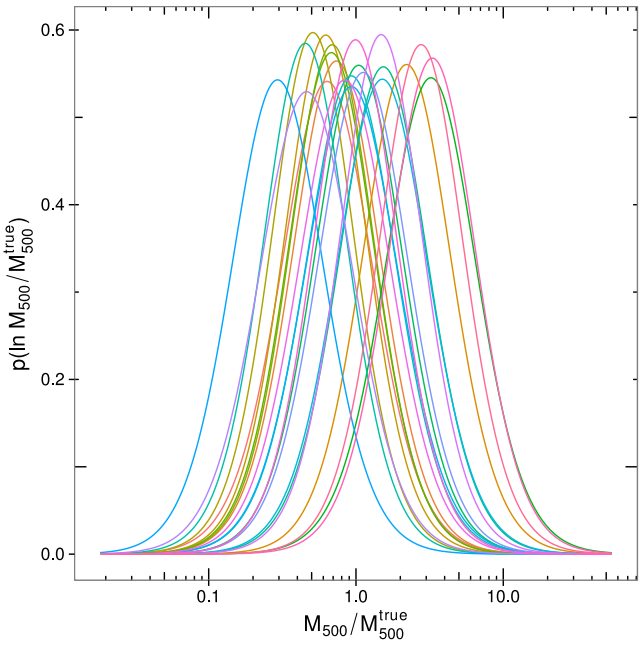

Figure A1. The posteriors for $M_{500}^{\rm true}$ using the generative model trained on a synthetic data set as described in Section 4.1 applied to 22 subsequent proxy observations. For each observation the inferred values of $M_{500}^{\rm true}$ have been normalized by the actual value of $M_{500}^{\rm true}$ so that the posteriors can be overplotted. Typical uncertainties in the inferred $M_{500}^{\rm true}$ for the calibration and observational uncertainties of this synthetic data set are approximately 0.5 dex.

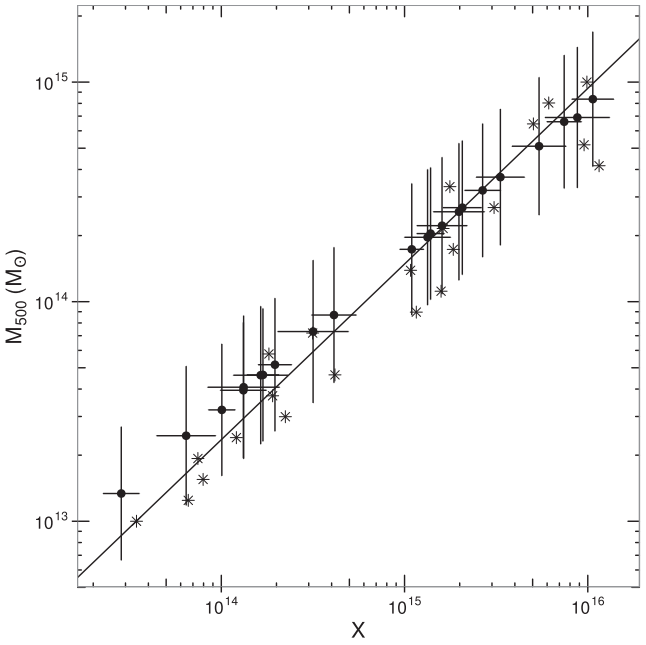

Figure A2. Predictions of $M_{500}^{\rm true}$ from observations of the mass proxy and associated errors for a test data set of 22 systems, after fitting the model to a similar training data set. Crosses show the measured proxy values and their errors (horizontally) and the mass estimates, with 1σ error regions, derived from the calculated posterior distribution for each system. Black stars give corresponding true masses and proxy values, and the solid line shows the mean linear relationship between $M_{500}^{\rm true}$ and $x_{\rm true}$ used to generate the data.

values (which have been subject to observational error) and the mass estimates derived from them via the posteriors from the generative model. These are compared to the values (stars) of $(x_{\rm true}, M_{500}^{\rm true})$.

Code that implements this generative model can be found under an MIT license at https://github.com/farr/galmassproxy. The model has been trained using the group/cluster data set from this paper, and can be used to generate a posterior mass distribution based on a measured value of one of the mass proxies we have calibrated.

This paper has been typeset from a TEX/LATEX file prepared by the author.